

Biophysical Journal, Volume 96

Supporting Material

**DNA Sequence-directed Organization of Chromatin: Structure-based
Computational Analysis of Nucleosome-binding Sequences**

Sreekala Balasubramanian, Fei Xu, and Wilma K. Olson

Supplementary Materials

**DNA Sequence-directed Organization of Chromatin:
Structure-based Analysis of Nucleosome-binding Sequences**

Sreekala Balasubramanian, Fei Xu, and Wilma K. Olson

Department of Chemistry & Chemical Biology

Rutgers, the State University of New Jersey

Wright-Rieman Laboratories

610 Taylor Road

Piscataway, New Jersey 08854-8087, USA

List of Supplementary Tables

- Table S1.** Protein-DNA complexes with resolution $\leq 2.5 \text{ \AA}$ used to generate knowledge-based potentials.
- Table S2.** Nucleosome-positioning sequences: (i) nucleosome-binding sequences from the mouse genome; (ii) anti-selection sequences from the mouse genome; (iii) synthetic SELEX sequences; (iv) high-resolution positioning sequences.
- Table S3.** Properties of DNA base-pair steps used in knowledge-based potentials: (i) number of base-pair steps included in derived ‘energy’ functions; (ii) average values and dispersion of Tilt, $\langle \theta_1 \rangle \pm (\sigma_{\theta_1})$, in protein-bound DNA dimers; (iii) average values and dispersion of Roll, $\langle \theta_2 \rangle \pm (\sigma_{\theta_2})$, in protein-bound DNA dimers; (iv) average values and dispersion of Twist, $\langle \theta_3 \rangle \pm (\sigma_{\theta_3})$, in protein-bound DNA dimers; (v) average values and dispersion of Shift, $\langle \theta_4 \rangle \pm (\sigma_{\theta_4})$, in protein-bound DNA dimers; (vi) average values and dispersion of Slide, $\langle \theta_5 \rangle \pm (\sigma_{\theta_5})$, in protein-bound DNA dimers; (vii) average values and dispersion of Rise, $\langle \theta_6 \rangle \pm (\sigma_{\theta_6})$, in protein-bound DNA dimers; (viii) relative base-pair flexibility V of protein-bound DNA dimers.
- Table S4.** Comparative ‘cost’ of nucleosomal deformation of individual base-pair steps with knowledge-based functions based on different subsets of observed protein-bound DNA conformations: (i) pyrimidine-purine (YR) base-pair steps; (ii) purine-purine and pyrimidine-pyrimidine (RR, YY) base-pair steps; (iii) purine-pyrimidine (RY) base-pair steps.
- Table S5.** Threading scores of known nucleosome-binding and nucleosome-resistant sequences on the central 60 base-pair steps of the 147-bp nucleosome core-particle structure (NDB_ID pd0287) (2) and an ideal 61-bp superhelical template with the same average global structure.

Table S1. Protein-DNA complexes with resolution ≤ 2.5 Å used to generate knowledge-based potentials.

NDB_ID	PDB_ID	Structure Description	Resol. (Å)	Literature citation
PD0002	1A73	INTRON-ENCODED ENDONUCLEASE I-PPOI/DNA COMPLEX	1.8	Flick <i>et al.</i> , 1998
PD0003	1CRX	CRE RECOMBINASE/DNA COMPLEX	2.4	Guo <i>et al.</i> , 1997
PD0006	3PVI	RESTRICTION ENDONUCLEASE PVUII/DNA COMPLEX	1.59	Horton <i>et al.</i> , 1998b
PD0007	9ANT	ANTENNAPEDIA HOMEODOMAIN/DNA COMPLEX	2.4	Fraenkel & Pabo, 1998
PD0008	1A6Y	REVERBA ORPHAN NUCLEAR RECEPTOR/DNA COMPLEX	2.3	Zhao <i>et al.</i> , 1998
PD0010	1BGB	ECORV ENDONUCLEASE/DNA COMPLEX	2.0	Horton & Perona, 1998a
PD0011	2PVI	RESTRICTION ENDONUCLEASE/DNA COMPLEX	1.76	Horton <i>et al.</i> , 1998a
PD0012	1BNZ	SSO7D HYPERTHERMOPHILE PROTEIN/DNA COMPLEX	2.0	Gao <i>et al.</i> , 1998
PD0013	1RV5	ECORV ENDONUCLEASE/DNA COMPLEX	2.1	Horton & Perona, 1998b
PD0016	3HDD	ENGRAILED HOMEODOMAIN/DNA COMPLEX	2.2	Fraenkel <i>et al.</i> , 1998
PD0020	1BC8	SAP-1 ETS DOMAIN/DNA COMPLEX	1.9	Mo <i>et al.</i> , 1998
PD0024	1B3T	EBNA-1 NUCLEAR PROTEIN/DNA COMPLEX	2.2	Bochkarev <i>et al.</i> , 1998
PD0027	1BC7	SERUM RESPONSE FACTOR ACCESSORY PROTEIN 1A (SAP-1)/DNA COMPLEX	2.0	Mo <i>et al.</i> , 1998
PD0028	1BG1	TRANSCRIPTION FACTOR STAT3B/DNA COMPLEX	2.2	Becker <i>et al.</i> , 1998
PD0029	2BAM	PROTEIN (ENDONUCLEASE BAMHI)/DNA COMPLEX	2.0	Viadiu & Aggarwal, 1998
PD0030	2KTQ	LARGE FRAGMENT OF DNA POLYMERASE I/DNA COMPLEX	2.3	Li <i>et al.</i> , 1998b
PD0031	3BAM	PROTEIN (RESTRICTION ENDONUCLEASE BAMHI) (3.1.21.4)/DNA COMPLEX	1.8	Viadiu, & Aggarwal, 1998
PD0032	3KTQ	LARGE FRAGMENT OF DNA POLYMERASE I/DNA COMPLEX	2.3	Li <i>et al.</i> , 1998b
PD0033	4KTQ	LARGE FRAGMENT OF DNA POLYMERASE I/DNA COMPLEX	2.5	Li <i>et al.</i> , 1998b
PD0035	1BDT	WILD TYPE GENE-REGULATING PROTEIN ARC/DNA COMPLEX	2.5	Schildbach <i>et al.</i> , 1999
PD0037	1B94	RESTRICTION ENDONUCLEASE ECORV/DNA COMPLEX	1.9	Thomas <i>et al.</i> , 1999
PD0042	1B8I	IV/HOMEBOX PROTEIN EXTRADENTICLE/DNA COMPLEX PRE-TRANSITION STATE ECO RI	2.4	Passner <i>et al.</i> , 1999
PD0049	1CKQ	ENDONUCLEASE/DNA COMPLEX	1.85	Horvath <i>et al.</i> , unpublished
PD0050	6PAX	HOMEBOX PROTEIN PAX-6/DNA COMPLEX	2.5	Xu <i>et al.</i> , 1999
PD0051	1CKT	HIGH MOBILITY GROUP PROTEIN HMG1/DNA COMPLEX	2.5	Ohndorf <i>et al.</i> , 1999

NDB_ID	PDB_ID	Structure Description	Resol. (Å)	Literature citation
PD0052	1SSP	URACIL-DNA GLYCOSYLASE/DNA COMPLEX	1.9	Parikh <i>et al.</i> , 1998
PD0054	1CL8	ENDONUCLEASE/DNA COMPLEX	1.8	Horvath <i>et al.</i> , unpublished
PD0055	1QPS	ENDONUCLEASE ECORI/DNA COMPLEX	2.5	Horvath <i>et al.</i> , unpublished
PD0056	1QPZ	PURINE NUCLEOTIDE SYNTHESIS REPRESSOR/DNA COMPLEX	2.5	Glasfeld <i>et al.</i> , 1999
PD0062	1QRH	RESTRICTION ENDONUCLEASE ECO-RI/DNA COMPLEX	2.5	Choi <i>et al.</i> , 2002
PD0065	1QSY	KLENOW FRAGMENT OF DNA POLYMERASE I/DNA COMPLEX	2.3	Li <i>et al.</i> , 1999
PD0066	1QSS	KLENOW FRAGMENT OF DNA POLYMERASE I/DNA COMPLEX	2.3	Li <i>et al.</i> , 1999
PD0067	1QTM	DNA POLYMERASE I (E.C.2.7.7.7)/DNA COMPLEX	2.3	Li <i>et al.</i> , 1999
PD0068	1QUM	ENDONUCLEASE IV (E.C.3.1.21.2)/DNA COMPLEX	1.6	Hosfield <i>et al.</i> , 1999
PD0070	1D3U	TATA-BINDING PROTEIN/DNA COMPLEX	2.4	Littlefield <i>et al.</i> , 1999
PD0071	1BY4	PROTEIN RETINOIC ACID RECEPTOR/DNA COMPLEX	2.1	Zhao <i>et al.</i> , 2000
PD0073	3HTS	HEAT SHOCK TRANSCRIPTION FACTOR/DNA COMPLEX	1.8	Littlefield & Nelson, 1999
PD0075	1B72	PBX1, HOMEBOX PROTEIN HOX-B1/DNA TERNARY COMPLEX	2.4	Piper <i>et al.</i> , 1999
PD0076	2IRF	INTERFERON REGULATORY FACTOR 2/DNA COMPLEX	2.2	Fujii <i>et al.</i> , 1999
PD0088	1BF4	PROTEIN (CHROMOSOMAL PROTEIN SSO7D)/DNA COMPLEX	1.6	Gao <i>et al.</i> , 1998
PD0089	1HWT	PROTEIN (HEME ACTIVATOR PROTEIN)/DNA COMPLEX	2.5	King <i>et al.</i> , 1999a
PD0090	2HAP	HEME ACTIVATOR PROTEIN HAP1-18/DNA COMPLEX	2.5	King <i>et al.</i> , 1999b
PD0096	1CYQ	INTRON-ENCODED HOMING ENDONUCLEASE I-PPOI/DNA COMPLEX	1.9	Galburt <i>et al.</i> , 1999
PD0099	1DIZ	3-METHYLADENINE DNA GLYCOSYLASE II/DNA COMPLEX	2.5	Hollis <i>et al.</i> , 2000
PD0101	1DFM	ENDONUCLEASE BGLII/DNA COMPLEX	1.5	Lukacs <i>et al.</i> , 2000
PD0108	1DMU	BGLI RESTRICTION ENDONUCLEASE/DNA COMPLEX	2.2	Newman <i>et al.</i> , 1998
PD0110	1QRV	HIGH MOBILITY GROUP PROTEIN D/DNA COMPLEX	2.2	Murphy IV <i>et al.</i> , 1999
PD0111	1DP7	MHC CLASS II TRANSCRIPTION FACTOR HRFX1/DNA COMPLEX	1.5	Gajiwala <i>et al.</i> , 2000
PD0115	1DSZ	RETINOIC ACID RECEPTOR RXR α & RAR α /DNA COMPLEX	1.7	Rastinejad <i>et al.</i> , 2000
PD0116	1DUX	ETS-DOMAIN PROTEIN ELK-1/DNA COMPLEX	2.1	Mo <i>et al.</i> , 2000
PD0117	1EBM	8-OXOGUANINE DNA GLYCOSYLASE/DNA COMPLEX	2.1	Bruner <i>et al.</i> , 2000
PD0119	1CA5	CHROMOSOMAL PROTEIN SAC7D/DNA COMPLEX	2.2	Su <i>et al.</i> , 2000
PD0120	1CA6	CHROMOSOMAL PROTEIN SAC7D/DNA COMPLEX	2.2	Su <i>et al.</i> , 2000
PD0121	1EGW	MADS BOX TRANSCRIPTION ENHANCER FACTOR 2/DNA COMPLEX	1.5	Santelli & Richmond, 2000
PD0122	1QPI	TETRACYCLINE REPRESSOR/DNA COMPLEX	2.5	Orth <i>et al.</i> , 2000

NDB_ID	PDB_ID	Structure Description	Resol. (Å)	Literature citation
PD0126	1QAI	REVERSE TRANSCRIPTASE/DNA COMPLEX	2.3	Najmudin <i>et al.</i> , 2000
PD0127	1EMH	URACIL-DNA GLYCOSYLASE/DNA COMPLEX	1.8	Parikh <i>et al.</i> , 2000
PD0131	1EON	TYPE II RESTRICTION ENZYME ECORV/DNA COMPLEX	1.6	Horton <i>et al.</i> , 2000
PD0132	1EEO	TYPE II RESTRICTION ENZYME ECORV/DNA COMPLEX	2.2	Horton & Perona, 2000
PD0139	1ESG	TYPE II RESTRICTION ENZYME BAMH/DNA COMPLEX	1.9	Viadiu & Aggarwal, 2000
PD0141	1EWN	3-METHYL-ADENINE DNA GLYCOSYLASE/DNA COMPLEX	2.1	Lau <i>et al.</i> , 2000
PD0142	1EWQ	DNA MISMATCH REPAIR PROTEIN MUTS/DNA COMPLEX	2.2	Obmolova <i>et al.</i> , 2000
PD0147	1EYU	TYPE II RESTRICTION ENZYME PVUII/DNA COMPLEX	1.8	Horton & Cheng, 2000
PD0151	1D02	RESTRICTION ENDONUCLEASE/DNA COMPLEX	1.7	Deibert <i>et al.</i> , 1999
PD0152	1D1U	MOLONEY MURINE LEUKEMIA VIRUS REVERSE TRANSCRIPTASE/DNA COMPLEX	2.3	Cote <i>et al.</i> , 2000
PD0153	1D2I	RESTRICTION ENDONUCLEASE BGLI/DNA COMPLEX	1.7	Lukacs <i>et al.</i> , 2000
PD0154	1QN4	TRANSCRIPTION INITIATION FACTOR TFIID-1/DNA COMPLEX	1.9	Patikoglou <i>et al.</i> , 1999
PD0160	1QNA	TRANSCRIPTION INITIATION FACTOR TFIID-1/DNA COMPLEX	1.8	Patikoglou <i>et al.</i> , 1999
PD0173	1MJ2	METHIONINE REPRESSOR MUTANT (Q44K)/DNA COMPLEX	2.4	Garvie & Phillips, 2000
PD0180	1GD2	TRANSCRIPTION FACTOR PAPI/DNA COMPLEX	2.0	Fujii <i>et al.</i> , 2000
PD0194	1FJX	HHAI METHYLTRANSFERASE MUTANT	2.3	Vilkaitis <i>et al.</i> , 2000
PDE001	1ERI	ECO RI ENDONUCLEASE (E.C.3.1.21.4)/DNA COMPLEX	2.5	Kim <i>et al.</i> , 1990
PDE005	1DNK	(E.C.3.1.21.1)/DNA COMPLEX	2.3	Weston <i>et al.</i> , 1992
PDE006	2DNJ	(E.C.3.1.21.1)/DNA COMPLEX	2.0	Lahm & Suck, 1991
PDE009	1HCR	HIN RECOMBINASE/DNA COMPLEX	2.3	Feng <i>et al.</i> , 1994
PDE0125	1BPY	DNA POLYMERASE β /DNA COMPLEX	2.2	Sawaya <i>et al.</i> , 1997
PDE0126	1BPX	DNA POLYMERASE β /DNA COMPLEX	2.4	Sawaya <i>et al.</i> , 1997
PDE0128	1TC3	TC3 TRANSPOSASE/DNA COMPLEX	2.5	Van Pouderoyen <i>et al.</i> , 1997
PDE0131	2BDP	DNA POLYMERASE I/DNA COMPLEX	1.8	Kiefer <i>et al.</i> , 1998
PDE0132	3BDP	DNA POLYMERASE I/DNA COMPLEX	1.9	Kiefer <i>et al.</i> , 1998
PDE0133	4BDP	DNA POLYMERASE I/DNA COMPLEX	1.8	Kiefer <i>et al.</i> , 1998
PDE0135	1T7P	DNA POLYMERASE, THIOREDOXIN/DNA COMPLEX	2.2	Doublie <i>et al.</i> , 1998
PDE014	1RVA	ECO RV/DNA COMPLEX	2.0	Kostrewa & Winkler, 1995
PDE0143	1A35	HUMAN TOPOISOMERASE I/DNA COMPLEX	2.5	Redinbo <i>et al.</i> , 1998

NDB_ID	PDB_ID	Structure Description	Resol. (Å)	Literature citation
PDE0145	1BSU	ENDONUCLEASE ECORV (E.C.3.1.21.4)/DNA COMPLEX	2.0	Martin <i>et al.</i> , 1999
PDE020	1BHM	BAMHI (E.C.3.1.21.4)/DNA COMPLEX	2.2	Newman <i>et al.</i> , 1995
PDE025	1FJL	SEGMENTATION PROTEIN PAIRED/DNA COMPLEX	2.0	Wilson <i>et al.</i> , 1995
PDE139	1AZ0	ECORV ENDONUCLEASE (E.C.3.1.21.4)/DNA COMPLEX	2.0	Perona & Martin, 1997
PDE141	6MHT	CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI (E.C.2.1.1.73)/DNA COMPLEX	2.05	Kumar <i>et al.</i> , 1997
PDR001	3CRO	434 CRO/DNA COMPLEX	2.5	Mondragon & Harrison, 1991
PDR004	2OR1	434 REPRESSOR/DNA COMPLEX	2.5	Aggarwal <i>et al.</i> , 1988
PDR010	1LMB	LAMBDA REPRESSOR/DNA COMPLEX	1.8	Beamer & Pabo, 1992
PDR011	1RPE	434 REPRESSOR/DNA COMPLEX	2.5	Shimon & Harrison, 1993
PDR013	1TRR	TRP REPRESSOR/DNA COMPLEX	2.4	Lawson & Carey, 1993
PDR015	1PER	434 REPRESSOR/DNA COMPLEX	2.5	Rodgers & Harrison, 1993
PDR018	1PDN	PRD PAIRED/DNA COMPLEX	2.5	Xu <i>et al.</i> , 1995
PDR021	2NLL	RETINOIC ACID RECEPTOR, THYROID HORMONE RECEPTOR/DNA COMPLEX	1.9	Rastinejad <i>et al.</i> , 1995
PDR022	1TSR	P53 TUMOR SUPPRESSOR/DNA COMPLEX	2.2	Cho <i>et al.</i> , 1994
PDR023	1BER	CAP/DNA COMPLEX	2.5	Parkinson <i>et al.</i> , 1996
PDR031	1AIS	TATA-BINDING PROTEIN/DNA COMPLEX	2.1	Kosa <i>et al.</i> , 1997
PDR032	1A3Q	HUMAN NF- κ B P52 BOUND TO DNA	2.1	Cramer <i>et al.</i> , 1997
PDR034	1AU7	PIT-1 MUTANT/DNA COMPLEX	2.3	Jacobson <i>et al.</i> , 1997
PDR036	1MNM	α -2 TRANSCRIPTIONAL REPRESSOR/DNA COMPLEX	2.2	Tan & Richmond, 1998
PDR047	1AZP	HYPERTHERMOPHILE CHROMOSOMAL PROTEIN SAC7D/DNA COMPLEX	1.6	Robinson <i>et al.</i> , 1998
PDR048	1AZQ	HYPERTHERMOPHILE CHROMOSOMAL PROTEIN SAC7D/DNA COMPLEX	1.9	Robinson <i>et al.</i> , 1998
PDR049	1AKH	MAT A1/ α 2/DNA TERNARY COMPLEX	2.5	Li <i>et al.</i> , 1998a
PDR051	2RAM	TRANSCRIPTION FACTOR NF- κ B P65/DNA COMPLEX	2.4	Chen <i>et al.</i> , 1998
PDR056	1BL0	MULTIPLE ANTIBIOTIC RESISTANCE PROTEIN/DNA COMPLEX	2.3	Rhee <i>et al.</i> , 1998
PDR056	1HCQ	ESTROGEN RECEPTOR/DNA COMPLEX	2.4	Schwabe <i>et al.</i> , 1993
PDT012	1YTB	TATA BINDING PROTEIN (TBP)/DNA COMPLEX	1.8	Kim <i>et al.</i> , 1993
PDT013	-	HNF-3/FORK HEAD DNA-RECOGNITION MOTIF/DNA COMPLEX	2.5	Clark <i>et al.</i> , 1993

NDB_ID	PDB_ID	Structure Description	Resol. (Å)	Literature citation
PDT015	1NFK	NUCLEAR FACTOR κ B (NF- κ B)/DNA COMPLEX	2.3	Ghosh <i>et al.</i> , 1995
PDT028	1YRN	MAT A1 HOMEODOMAIN, MAT α 2 HOMEODOMAIN/DNA COMPLEX	2.5	Li <i>et al.</i> , 1995
PDT029	2DGC	GCN4/DNA COMPLEX	2.2	Keller <i>et al.</i> , 1995
PDT030	1LAT	GLUCOCORTICOID RECEPTOR MUTANT/DNA COMPLEX	1.9	Gewirth & Sigler, 1995
PDT031	–	EVEN-SKIPPED HOMEODOMAIN/DNA COMPLEX	2.0	Hirsch & Aggarwal, 1995
PDT033	1PUE	TRANSCRIPTION FACTOR PU.1/DNA COMPLEX	2.1	Kodandapani <i>et al.</i> , 1996
PDT034	1CDW	TATA BINDING PROTEIN (TBP)/DNA COMPLEX	1.9	Nikolov <i>et al.</i> , 1996
PDT035	1IGN	RAP1/DNA COMPLEX	2.2	Koenig <i>et al.</i> , 1996
PDT036	1YTF	TRANSCRIPTION FACTOR IIA (TFIIA)/DNA COMPLEX	2.5	Tan <i>et al.</i> , 1996
PDT038	1UBD	YY1 ZINC FINGER DOMAIN MUTANT/DNA COMPLEX	2.5	Houbaviy <i>et al.</i> , 1996
PDT039	1AAAY	ZIF268 ZINC FINGER PEPTIDE/DNA COMPLEX	1.6	Elrod-Erickson <i>et al.</i> , 1996
PDT040	1IHF	INTEGRATION HOST FACTOR/DNA COMPLEX	2.2	Rice <i>et al.</i> , 1996
PDT043	2HDD	ENGRAILED HOMEODOMAIN Q50K VARIANT/DNA COMPLEX	1.9	Tucker-Kellogg <i>et al.</i> , 1997
PDT044	1ZME	PROLINE UTILIZATION TRANSCRIPTION ACTIVATOR (PUT3)/DNA COMPLEX	2.5	Swaminathan <i>et al.</i> , 1997
PDT045	1XBR	T DOMAIN/DNA COMPLEX	2.5	Muller & Herrmann, 1997
PDT048	1AWC	GA BINDING PROTEIN α , GA BINDING PROTEIN β 1/DNA COMPLEX	2.1	Batchelor <i>et al.</i> , 1998
PDT049	2CGP	CATABOLITE GENE ACTIVATOR PROTEIN/DNA COMPLEX	2.2	Passner & Steitz, 1997
PDT055	1A1F	THREE-FINGER ZIF268 PEPTIDE/DNA COMPLEX	2.1	Elrod-Erickson <i>et al.</i> , 1998
PDT056	1A1G	THREE-FINGER ZIF268 PEPTIDE/DNA COMPLEX	1.9	Elrod-Erickson <i>et al.</i> , 1998
PDT057	1A1H	QGSR (THREE-FINGER ZIF268 VARIANT)/DNA COMPLEX	1.6	Elrod-Erickson <i>et al.</i> , 1998
PDT059	1A1I	ZINC FINGER/DNA COMPLEX	1.6	Elrod-Erickson <i>et al.</i> , 1998
PDT062	1AM9	STEROL REGULATORY ELEMENT BINDING(STRE)/DNA COMPLEX	2.3	Parraga <i>et al.</i> , 1998
PDTB41	1MEY	CONSENSUS ZINC FINGER/DNA COMPLEX	2.2	Kim & Berg, 1996
PDV001	2BOP	E2/DNA COMPLEX	1.7	Hegde <i>et al.</i> , 1992

Note: Multiple entries of *Eco* RV endonuclease reflect different structures solved to study the different functional aspects of the catalytic activity of the enzyme.

References to Table S1

- Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988) Recognition of a DNA operator by the repressor of Phage 434: A view at high resolution. *Science* **242**, 899-907.
- Batchelor, A.H., Piper, D.E., Charles De La Brousse, D., McKnight, S.L. & Wolberger, C. (1998) The structure of GABP α/β : an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science* **279**, 1037-1041.
- Beamer, L.J. & Pabo, C.O. (1992) Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.*, **227**, 177-196.
- Becker, S., Groner, B. & Muller, C.W. (1998) Three-dimensional structure of the Stat3 β homodimer bound to DNA. *Nature* **394**, 145-151.
- Bochkarev, A., Bochkareva, E., Edwards, A.M. & Frappier, L. (1998) The 2.2 Å structure of a permanganate-sensitive DNA site bound by the Epstein-Barr virus origin binding protein, EBNA1. *J. Mol. Biol.* **284**, 1273-1278.
- Bruner, S.D., Norman, D.P. & Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859-866.
- Chen, Y.Q., Ghosh, S. & Ghosh, S. (1998) A novel DNA recognition mode by the NF- κ B P65 homodimer. *Nature Struct. Biol.* **5**, 67-73.
- Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994) Crystal structure of a P53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **265**, 346-355.
- Choi, J., Kim, Y., Greene, P., Hager, P. & Rosenberg, J.M. (to be published) X-ray structure of the DNA-*Eco* RI endonuclease complexes with the ED144 & RK145 mutations.
- Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. (1993) Co-crystal structure of the HNF-3/Fork Head DNA-recognition motif resembles histone H5. *Nature* **364**, 412-420.
- Cote, M.L., Yohannan, S.J. & Georgiadis, M.M. (2000) Use of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase to facilitate crystallization and analysis of a pseudo-16-mer DNA molecule containing G-A mispairs. *Acta Crystallogr., Sect. D* **56**, 1120-1131.
- Cramer, P., Larson, C.J., Verdine, G.L. & Muller, C.W. (1997) Structure of the human NF- κ B p52 homodimer-DNA complex at 2.1 Å resolution. *EMBO J.* **16**, 7078-7090.
- Deibert, M., Grazulis, S., Janulaitis, A., Siksny, V. & Huber, V. (1999) Crystal structure of Mun I restriction endonuclease in complex with cognate DNA at 1.7 Å Resolution. *EMBO J.* **18**, 5805-5816.
- Double, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251-258.
- Elrod-Erickson, M., Benson, T.E. & Pabo, C.O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464.
- Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180.
- Feng, J.A., Johnson, R.C. & Dickerson, R.E. (1994) Hin recombinase bound to DNA: The origin of specificity in major and minor groove interactions. *Science* **263**, 348-355.
- Flick, K.E., Jurica, M.S., Monnat Jr., R.J. & Stoddard, B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **394**, 96-101.
- Fraenkel, E. & Pabo, C.O. (1998) Comparison of X-ray and NMR structures for the *Antennapedia* homeodomain-DNA complex. *Nature Struct. Biol.* **5**, 692-697.

- Fraenkel, E., Rould, M.A., Chambers, K.A. & Pabo, C.O. (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.* **284**, 351-361.
- Fujii, Y., Shimizu, T., Kusumoto, M., Kyogoku, Y., Taniguchi, T. & Hakoshima, T. (1999) Crystal structure of an IRF/DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequence. *EMBO J.* **18**, 5028-5041.
- Fujii, Y., Shimizu, T., Toda, T., Yanagida, M. & Hakoshima, T. (2000) Structural basis for the DNA recognition diversity of bZIP transcription factors revealed by the crystal structure of Pap1/DNA complex. *Nature Struct. Biol.* **7**, 889-893.
- Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B. & Burley, S.K. (2000) Structure of the winged helix-protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921.
- Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat Jr., R.J. & Stoddard, B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nature Struct. Biol.* **6**, 1096-1099.
- Gao, Y.G., Su, S.Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B.S., Edmondson, S.P., Shriver, J.W. & Wang A. H. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7D bound to DNA. *Nature Struct. Biol.* **5**, 782-786.
- Garvie, C.W. & Phillips, S.E.V. (2000) Direct and indirect readout in mutant Met repressor-operator complexes. *Structure Fold. Des.* **8**, 905-914.
- Gewirth, D.T. & Sigler, P.B. (1995) The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex. *Nature Struct. Biol.* **2**, 386-394.
- Ghosh, G., Van Duyne, G., Ghosh, S. & Sigler, P.B. (1995) Structure of NF-κB p50 homodimer bound to a κB site. *Nature* **373**, 303-310.
- Glasfeld, A., Koehler, A.N., Schumacher, M.A. & Brennan, R.G. (1999) The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions. *J. Mol. Biol.* **291**, 347-361.
- Guo, F., Gopaul, D.N. & Van Duyne, G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-46.
- Hegde, R.S., Grossman, S.R., Laimins, L.A. & Sigler, P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505-512.
- Hirsch, J.A. & Aggarwal, A. K. (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: New perspectives on homeodomain specificity. *EMBO J.* **14**, 6280-6291.
- Hollis, T., Ichikawa, Y. & Ellenberger, T. (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J.* **19**, 758-766.
- Horton, J.R., Bonventre, J. & Cheng, X. (1998a) How is modification of the DNA substrate recognized by the *PvuII* restriction endonuclease? *J. Biol. Chem.* **379**, 451-458.
- Horton, J.R. & Cheng, X. (2000) *PvuII* endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **300**, 1049-1056.
- Horton, N.C., Connolly, B.A. & Perona, J.J. (2000) Inhibition of EcoRV endonuclease by deoxyribo-3'-S-phosphorothiolates: a high-resolution X-ray crystallographic study *J. Am. Chem. Soc.* **122**, 3314-3324.
- Horton, J.R., Nastro, H.G., Riggs, P.D., & Cheng, X. (1998b) Asp34 of *PvuII* endonuclease is directly involved in DNA minor groove recognition and indirectly involved in catalysis. *J. Mol. Biol.* **284**, 1491-1504.
- Horton, N.C. & Perona, J.J. (1998a) Recognition of flanking DNA sequences by EcoRV endonuclease involves alternative patterns of water-mediated contacts. *J. Biol. Chem.* **273**, 21721-21729.
- Horton, N.C. & Perona, J.J. (1998b) Role of protein-induced bending in the specificity of DNA recognition: crystal structure of EcoRV endonuclease complexed with d(AAAGAT) + d(ATCTT). *J. Mol. Biol.* **277**, 779-783.

- Horton, N.C. & Perona, J.J. (2000) Crystallographic snapshots along a protein-induced DNA-bending pathway. *Proc. Natl. Acad. Sci., USA* **97**, 5729-5734.
- Horvath, M.M., Choi, J., Kim, Y., Wilkosz, P. & Rosenberg, J.M. (to be published) The integration of recognition and cleavage: X-ray structures of pre- transition state and post-reactive DNA-*Eco* RI endonuclease complexes.
- Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P. & Tainer, J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **98**, 397-408.
- Houbaviy, H.B., Usheva, A., Shenk, T. & Burley, S.K. (1996) Cocystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci., USA* **93**, 13577-13582.
- Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G. & Aggarwal, A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.* **11**, 198 - 212.
- Keller, W., Koenig, P. & Richmond, T.J. (1995) Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *J. Mol. Biol.* **254**, 657-667.
- Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304-307.
- Kim, C.A. & Berg, J.M. (1996) A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature Struct. Biol.* **3**, 940-945.
- Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520.
- Kim, Y., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. (1990) Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307-1309.
- King, D.A., Zhang, L., Guarente, L. & Marmorstein, R. (1999a) Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein. *Nature Struct. Biol.* **6**, 64-71.
- King, D.A., Zhang, L., Guarente, L. & Marmorstein, R. (1999b) Structure of HAP1-18-DNA implicates direct allosteric effect of protein-DNA interactions on transcriptional activation. *Nature Struct. Biol.* **6**, 22-27.
- Kodandapani, R., Pio, F., Ni, C.Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A. & Ely, K.R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**, 456-460.
- Koenig, P., Giraldo, R., Chapman, L. & Rhodes, D. (1996) Crystal structure of the DNA binding domain of yeast RAP1 in complex with a telomeric DNA site. *Cell* **85**, 125-136.
- Kosa, P.F., Ghosh, G., DeDecker, B.S. & Sigler, P.B. (1997) The 2.1- Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci., USA* **94**, 6042-6047.
- Kostrewa, D. & Winkler, F.K. (1995) Mg²⁺ binding to the active site of *Eco*RV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* **34**, 683-696.
- Kumar, S., Horton, J.R., Jones, G.D., Walker, R.T., Roberts, R.J. & Cheng, X. (1997) DNA containing 4'-thio-2'-deoxycytidine inhibits methylation by HhaI methyltransferase. *Nucleic Acids Res.* **25**, 2773-2783.
- Lahm, A. & Suck, D. (1991) DNase I-induced DNA conformation. 2 Å structure of a DNase I-octamer complex. *J. Mol. Biol.* **222**, 645-667.
- Lau, A.Y., Wyatt, M.D., Glassner, B.J., Samson, L.D. & Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci., USA* **97**, 13573-13578.
- Lawson, C.L. & Carey, J. (1993) Tandem binding in crystals of a trp repressor/operator half-site complex. *Nature* **366**, 178-182.

- Li, T., Stark, M.R., Johnson, A.D. & Wolberger, C. (1995) Crystal structure of the MAT α 1/MAT α 2 homeodomain heterodimer bound to DNA. *Science* **270**, 262-269.
- Li, T., Jin, Y., Vershon, A.K. & Wolberger, C. (1998a) Crystal structure of the MAT α 1/MAT α 2 homeodomain heterodimer in complex with DNA containing an A-tract. *Nucleic Acids Res.* **26**, 5707-5718.
- Li, Y., Korolev, S. & Waksman, G. (1998b) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.
- Li, Y., Mitaxov, V., Waksman, G. (1999) Structure-based design of *Taq* DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci., USA* **96** 9491-9496.
- Littlefield, O., Korkhin, Y. & Sigler, P.B. (1999) The structural basis for the oriented assembly of a TBP/Tfb/promoter complex. *Proc. Natl. Acad. Sci., USA* **96**, 13668-13673.
- Littlefield, O. & Nelson, H.C.M. (1999) A new use for the 'wing' of the winged helix-turn-helix motif in the Hsf-DNA cocrystal. *Nature Struct. Biol.* **6**, 464-470.
- Lukacs, C.M., Kucera, R., Schildkraut, I. & Aggarwal, A.K. (2000) Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 Å resolution. *Nature Struct. Biol.* **7**, 134-140.
- Martin, A.M., Sam, M.D., Reich, N.O. & Perona, J.J. (1999) Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nature Struct. Biol.* **6**, 269-277.
- Mo, Y., Vaessen, B., Johnston, K. & Marmorstein, R. (1998) Structures of SAP-1 Bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by ETS proteins. *Mol. Cell* **2**, 201-212.
- Mo, Y., Vaessen, B., Johnston, K. & Marmorstein, R. (2000) Structure of the elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. *Nature Struct. Biol.* **7**, 292-297.
- Mondragon, A. & Harrison, S.C. (1991) The phage 434 Cro/OR1 complex at 2.5 Å resolution. *J. Mol. Biol.* **219**, 321-334.
- Muller, C.W. & Herrmann, B.G. (1997) Crystallographic structure of the T domain-DNA complex of the *Brachyury* transcription factor. *Nature* **389**, 884-888.
- Murphy IV, F.V., Sweet, R.M. & Churchill, M.E.A. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J.* **18**, 6610-6618.
- Najmudin, S., Cote, M.L., Sun, D., Yohannan, S., Montano, S.P., Gu, J. & Georgiadis, M.M. (2000) Crystal structures of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase complexed with nucleic acid: functional implications for template-primer binding to the fingers domain. *J. Mol. Biol.* **296**, 613-622.
- Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. & Phillips, S.E.V. (1998) Crystal structure of restriction endonuclease BglII bound to its interrupted DNA recognition sequence. *EMBO J.* **17**, 5466-5476.
- Newman, M., Strzelecka, T., Dorner, L.F., ISchildkraut, I. & Aggarwal, A.K. (1995) Structure of BamHI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269**, 656-663.
- Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G. & Burley, S.K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci., USA* **93**, 4862-4867.
- Obmolova, G., Ban, C., Hsieh, P. & Yang, W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-710.
- Ohndorf, U.M., Rould, M.A., He, Q., Pabo, C.O. & Lippard, S.J. (1999) Molecular basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708-712.

- Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Struct. Biol.* **7**, 215-219.
- Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E. & Tainer, J.A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214-5226.
- Parikh, S.S., Slupphaug, G., Krokan, H.E., Blackburn, G.M. & Tainer, J.A. (2000) Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci., USA* **97**, 5083-5088.
- Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.H. & Berman, H.M. (1996) Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* **260**, 395-408.
- Parraga, A., Bellolell, L., Ferre-D'Amare, A.R. & Burley, S.K. (1998) Co-crystal structure of sterol regulatory element binding protein 1A at 2.3 Å resolution. *Structure* **6**, 661-672.
- Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S. & Aggarwal, A.K. (1999) Structure of a DNA-bound ultrathorax-extradenticle homeodomain complex. *Nature* **397**, 714-719.
- Passner, J.M. & Steitz, T.A. (1997) The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proc. Natl. Acad. Sci., USA* **94**, 2843-2847.
- Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.-H., Kodadek, T. & Burley, S.K. (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217-3230.
- Perona, J.J. & Martin, A.M. (1997) Conformational transitions and structural deformability of EcoRV endonuclease revealed by crystallographic analysis. *J. Mol. Biol.* **273**, 207-225.
- Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., Wolberger, C. (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**, 587-597.
- Rastinejad, F., Perlmann, T., Evans, R.M. & Sigler, P.B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* **375**, 203-211.
- Rastinejad, F., Wagner, T., Zhao, Q. & Khorasanizadeh, S. (2000) Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J.* **19**, 1045-1054.
- Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J. & Hol, W.G. (1998) Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* **279**, 1504-1513.
- Rhee, S., Martin, R.G., Rosner, J.L. & Davies, D.R. (1998) A novel DNA-binding motif of MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl. Acad. Sci., USA* **95**, 10413-10418.
- Rice, P.A., Yang, S.W., Mizuuchi, K. & Nash, H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-1306.
- Robinson, H., Gao, Y.G., McCrary, B.S., Edmondson, S.P., Shriver, J.W. & Wang, A.H. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202-205.
- Rodgers, D.W. & Harrison, S.C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure* **1**, 227-240.
- Santelli, E. & Richmond, T.J. (2000) Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution. *J. Mol. Biol.* **297**, 437-449.
- Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J. & Pelletier, H. (1997) Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry* **36**, 11205-11215.

- Schildbach, J.F., Karzai, A.W., Raumann, B.E. & Sauer, R.T. (1999) Origins of DNA-binding specificity: role of protein contacts with the DNA backbone. *Proc. Natl. Acad. Sci., USA* **96**, 811-817.
- Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75**, 567-578.
- Shimon, L.J. & Harrison, S.C. (1993) The phage 434 OR2/R1-69 complex at 2.5 Å resolution. *J. Mol. Biol.* **232**, 826-838.
- Su, S., Gao, Y.G., Robinson, H., Liaw, Y.C., Edmondson, S.P., Shriver, J.W. & Wang, A.H.-J. (2000) Crystal structures of the chromosomal proteins Sso7D/Sac7D bound to DNA containing T-G mismatched base-pairs. *J. Mol. Biol.* **303**, 395-403.
- Swaminathan, K., Flynn, P., Reece, R.J. & Marmorstein, R. (1997) Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn₂Cys₆ binuclear cluster. *Nature Struct. Biol.* **4**, 751-759.
- Tan, S., Hunziker, Y., Sargent, D.J. & Richmond, T.J. (1996) Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127-134.
- Tan, S. & Richmond, T.J. (1998) Crystal structure of the yeast MAT α 2/MCM1/DNA ternary complex. *Nature* **391**, 660-666.
- Thomas, M.P., Brady, R.L., Halford, S.E., Sessions, R.B. & Baldwin, G.S. (1999) Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: a catalytic role for a main chain carbonyl group. *Nucleic Acids Res.* **27**, 3438-3445.
- Tucker-Kellogg, L., Rould, M.A., Chambers, K.A., Ades, S.E., Sauer, R.T. & Pabo, C.O. (1997) Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. *Structure* **5**, 1047-1054.
- Van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H.A. & Sixma, T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J.* **16**, 6044-6054.
- Viadiu, H. & Aggarwal, A.K. (1998) The role of metals in catalysis by the restriction endonuclease BamHI. *Nature Struct. Biol.* **5**, 910-916.
- Viadiu, H. & Aggarwal, A.K. (2000) Structure of BamHI bound to non-specific DNA: a model for DNA sliding. *Mol. Cell* **5**, 889-895.
- Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X. & Klimasauskas, S. (2000) Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. *J. Biol. Chem.* **275**, 38722-38730.
- Weston, S.A., Lahm, A. & Suck, D. (1992) X-ray structure of the DNase I-d(GGTATACC)₂ complex at 2.3 Å resolution. *J. Mol. Biol.* **226**, 1237-1256.
- Wilson, D.S., Guenther, B., Desplan, C. & Kuriyan, J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709-719.
- Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. & Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.* **13**, 1263-1275.
- Xu, W., Rould, M.A., Jun, S., Desplan, C. & Pabo, C.O. (1995) Crystal structure of a paired domain-DNA complex at 2.5 Ångstroms resolution reveals structural basis for Pax developmental mutations. *Cell* **80**, 639-650.
- Zhao, Q., Chasse, S.A., Devarakonda, S., Sierk, M.L., Ahvazi, B. & Rastinejad, F. (2000) Structural basis of RXR-DNA interactions. *J. Mol. Biol.* **296**, 509-520.
- Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M.A. & Rastinejad, F. (1998) Structural elements of an orphan nuclear receptor-DNA complex. *Mol. Cell* **1**, 849-861.

Table S2. Nucleosome-positioning sequences

(i) Nucleosome-binding sequences from the mouse genome
SET 1A A-tracts 1 (90%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phen4	122 bp	TAGGACCTGG AGTATGGCGA GAAAAC TGAA AATCACAGAA AATGAGAAAT ACACACTTTA GGATGTGAAA TATGGCGAGG AAAACTGAAA AAGGTGGAAA ATTCAGAAAT GTCCACTGTA GG
phen8	127 bp	AGGACCTGGA ACATGGTGAG AAAACTGAAA ATCACAGAAA ATGAGAAATA GACACTTTAG GACGTGAAAT ATGACGAGGA AAACAGAAAA AGTTGGAAAA TTTAGAAATG TCTAACGTAG GACGTGG
phen12	123 bp	TGAAGGACCT GGAATATGGC GACGGAAAAC TGAAAATCAC GGAAAATGAG AAATACACAC TTTACGACGT GAAATATGGC GAGGAAAAC TATAAAGGTG GAATATTTAG AAACGTCCAC TGT
phen14	125 bp	AGGGACATGG AATATGGAGA GAAAAC TGAA AATCACGGAA AATGAAAAAT ACACACTTTA GGACGTGAAA TATAGCGAGG AAAACTGAAA AAGGTGGAAAT ATTTAGAAAT GTCCACTGTA GGACG
phen17	120 bp	CCGGAATATG GCGAGAAAAC TGAAAATCAC GTAAAATGAG AAATACACAC TTTAGGACGT GAGATATCGC GAGGAAACCT GAAAAAGGTG GAAAATTTAG AAATGTCACA GTAGGACGTG
phen18	125 bp	TGACCTGGAA TATGGCGAGA AACCTGAAAA TCACGCAAAA TGAGAAATAC ACACTTTAGG ACATGAAATA TGGTGAGGAA AATTGAAAAA GGTGGAATAT TAAGAAATGT CCACTGTAGG ACGTG
phen21	129 bp	TGAAGGACCT GGAATATGGC GAGAGAACTG AAAATCACCG AAAATGAGAA ATACACACTT TAGGACGTGA AATATGGCGA GGAAAAC TGA AAAAGATGGA AAATTTAGAA ATATCCACTG TAGGACGTG
phen22	118 bp	AATGTGGCGA GAAAAGTGAA AATCACGGAA AATGAGAAAT AAACACTTTA GGAAGTGAAA TATGGCGAGG AAAACTGAAA AGGATGGAAA ATTTAGAAAT GTCCACTGTA GGACGTGG
phen23	125 bp	GAGGACCTGG AATATGGTGA GAAGACTGAA AATCACGGAA AATGAGAAAT ACACACTTTT GGACGTGAAA TATGGCGAGG AAAACTGAAA AAGGTGGAAA ATTTAGAAAT GTCCACTGTA GGACG
phen24	111 bp	AGAGAAAACC GAAAATCACG GAAAATGAGA AATACGCACT TTAGGACGTG AAATATGGCG AGGAAAAC TGA AAAAAGGTGG AAAATTTAGG GATGTCCACT GTAGGACGTG G
phen26	127 bp	AGGACCTGGA ACATGGCGAG AAAACTGAAG ATCACGGAAA ATGAGAAATA CACACTTTAG GGCGTGAAAT ATGACGAGGA AAACTGAAAA AAGGTGGAGAA TTTAGAAATG TCCACTGTAG GACGTGG

phcn29	120 bp	GGAATATGGC GAGAAAACAG AAAATCACGG GAAATGAGAA ATACACACTT TAGGACGTGA AATATAGCGA GGGGAACTGA AAAAGGTGGA AAATTTAGAA ATGTCCGCTG TAGGACGTGG
phcn39	128 bp	AGAGGACCTG GAATATGGCG AGAAAACCTGA AAATCACGGA AAATGAGAAA TACACACTTT AGGACATGAA ATATGGCGAG GAAAACCTGAA AAAGGTGGA AATTGAGAAA TGCCACTGTA GGACGTGG
phcn43	119 bp	AGGACCTGGA ATATGGCGAG AAAACTGAAA ATCACGGAAA ATGAGAAACA CGCGCTTAAG GACATGAAAT ATGGCGAGAA AAACTGAAAA AGGTGGAATA TATAGAAATG TCCACTGTA
phcn46	126 bp	ATGCACACTG TAGGACCTGG AATATGGCGA GAAAACCTGAA AATTAAGGAA AATGAGAAAT ATACACTTTA GGACGTGAAA TATGGCGAGG AGGACTGAAA GAGGTGGAAA ATTTAGATAC GTCCAC
phcn47	128 bp	AAGGACCTGG AATATGACGA GAAAACCTGAA AATCACGGAA AATGAGAGAT ACACACTTTA GGACGTGAAA TATGGCGAAG AAAACTGAAA AGGTCGGAAA ATTTAGAAAT GTCCACCGTA GACGTGGA
phcn50	109 bp	AGAGAACTGA AAATCACCGA AAATGAGAAA TACACGCTTT AGGACGTGAA ATATGGCGAG GAAAACCTGAA AAAGTGGAAA ATTTAGAAAT GTCCACTGTA GGACGTGGT
phwn12	126 bp	GACCTGGAAA ATGGCGAGAA AACTGAAAAT CACGAAAAT GTGAAATACA CACTTTAGGA CATGAAATAT GGCGAGGAAA ATTGAAAAAG TTGGATAATT TAGAAATGTC CACTGTAGGA CGTGA

SET 1B A-tracts 2 (81%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phcn2	126 bp	TAGGACGTGG AATATGGCAA GAAAAC TGAA AATCATGGAA AATGATAAGC ATCCACTTGA CGACTTGAAA AATGACGAAA TCACTAAAAA ACGTGAAAAA TGAGAAATGC ACACTGGAGG ACCTGG
phcn10	125 bp	AGAACGTGGA ATAAGGCAAG GAAACGGAAA ATCATGGAAA ATGAGGAACA TCCAATTGAC GACTGGAAAA ATGACGAAAT CACTAAAAAA CGTGAAAAAT GAGAAATGGA CACTGAAGGA CCTGG
phcn13	119 bp [†]	AGGACGTGAA AATTGGCGAGA AACTGGAAAAG GTGGAATATT TAGAGATGTG CACTGTAGGA CGTGGAAATAT GGCAAGAAAAG TGAAAAGCAT GGAAAATGAG AAACATCCAC TTGACGATT
phcn20	126 bp	TAGGACGTGG AATATGGCAA GAAAAC TGAA AATCATGGAA AATGAGAAAC ATCCACTTGA TGACTTGAAA AATGACGAAA TCACTAAAAA ACGTGAAAAA TGAGAAATGC ACACTGAAGG ACCTGA
phcn30	126 bp	TGACGACTTG AAAAATGACG AGATCACTAA AAAATGTGAA AAATGAGAAA TGCACACTGA AGGAGCTGGA ATATGGTGGG AAAACTGGAA ATTACGGAAA ATGAGAAATA CACACTTTAG GACGTG
phcn42	126 bp	AGGACGTGGG AGTATGGCAA GAAAAC TGAA AGTCATGGAA AGTGAGAAAC ATCCACTTGA TGACTTGAAA AATGACGAAA TCACTGAAAG ACGTGAAAAA TGAGAAATGCA CACTGTAGGA CATGGA
phwn8	119 bp	ATGGCAAGAA AACTGAAAAT CATGGAAAAT GAGAAACATC CACTCGACGA CTTGAAAAGAT GACGAAATCA CTAGAAAACG TGAAAAATGA GAAATGCTCA CTGAAGGACC TGAATATGG
phwn26	124 bp	TAGGACGTGG AATATGGCAA GAAATCTGAA AATCATGGAA AATGAGAAGC ATCCACTTGA TGACTTGAAA ATGACGAAAT CACTAAAAAA ACGTGAAAAAT GAGAAATGCA CACCGTAGAC CTGG
phwn27	121 bp	AGGACGTGGA ATACGGCAAG AAAACTGAAA ATCATGGAAA ATGAGAAACA TCCACTTGAC GACTTGAAAA ATGACGAAAT CACTAAAAAA ACGTGAAAAAT GAGAGATACA CACTGAAGGC C

[†] Chain length without space reported in sequence dataset of M. Kubista.

SET 1C A-tracts 3 (72%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phen3	123 bp	ACGTGAAAAA TGAGAAATGC AACTGAAGG ACCTGAAATA TGGCGAGAAA ACTGAAAATC ACGGAAAATG AGAAATACAC ACTTTAGGAC GTGAAAATATG GCGAGGAAAA CTGACAAAGG CGG
phen7	118 bp	AGAAAAGTGA AAATCACGGA AAATTAGAAA TACACACTTT AGGACGTGAA ATATGGCGAG GAAAATGGAA AAAGGTGGAA AATTTAGCAG TGTCCACTGT AGGAGGTGGA ATATGGCA
phen9	136 bp	AGGACGTGCA ATAAGGCAAG AAAACTGAAA ATCATGGAAA ATGAGAAACA TCCACTTGAC GACTTGAAAA ATGACAAAAT CACTAAAATA CGTGAAAAAT GAGAGATGCA CCCTGAAGGA CCTGGAATAT GGCGAG
phen11	124 bp	ATGGCTAGAA AACTGAAAAT CATGGAAAAA GAGAAACATC CACTTGACGA CTTGAAAAAT GTCGAAATCA CTAAAAAACG TGAAAAATGA GAAATGCACA CTGAGGGACC TGGAAATATGG CGAG
phen19	129 bp	ATGGCGAGAA AACTGAAGTT CACGGAAAAT GGAGAAATAC ACACTTTAGG ATGCAGTGAA ATATTGAGCG AAGGAAAAC TAAAAAGGTG GAAAAATTTAG AAATGTCCAC TGTAGGACGA GGAATATGG
phen28	122 bp	AGAAAAGTGA AAATCACGGA AAATGAGGAA TACACACTTT AGGACGTGAA ATATGGCGAG GAAAAGTGA AAAGTGGGAA AATTTAGAAA TGTCCACTGT AGGATCGTGG AATATAGCAG GC
phen37	117 bp	AGAACACTGA AAATCACGGA AAATGAGAAA TACACACTTT AGGACGTGAA ATATGGCGAG GAAAAGTAAA AAAGGTGGAA AATTTAGAAA TGTCCACTGT AGGACGTGGA ATATGGC
phen44	123 bp	ATGGCGAGAA AACTGAAAAT CACGGAAAAT GAAAAATACA CACTTCAGGA CGTGAAAATAT GGCGAGGAAA ACTGAAAAAG GTGGAAAATT TAGAAATGTC CACTGTAGGA CGTGAATAT GGC
phwn11	123 bp	ATGGCGAGAA AACTGCAAAT CACGGAGAAT GAGAAATACA CACTTTAGGA CGTGAAGTAT GGCGAGGAAA ACTGAAAAAG GTGGAAAATT TAGAAATGTC CGCTGTAGGA CGTGAATAT GGC
phwn29	136 bp	AGGACGTGCA ATAAGGCAAG AAAACTGAAA ATCATGGAAA ATGAGAAACA TCCACTTGAC GACTTGAAAA ATGACAAAAT CACTAAAATA CGTGAAAAAT GAGAAATGCA CCCTGAAGGA CCTGGAATAT GGCGAG
phwn36	120 bp	TGAAAAATTA GAAACATCCA CTTGATGACT TGAATAATGA CGAAATCACT AAAAAACGTG AAAGATGAGA AATGCACACT GAAGGACCTG GAATATGGCG AGAAAAGTGA AAATCACGGA

SET 2 TG/CTG RUNS (47%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phen31	141 bp	GCATGCATGC ATACACGCAT GTATACGAAC ATGTACACAC ATACTCATGA ACACGCATAC ACGCGCACAT ATGCATACAT GCTTGTATAC ACACATGTAT ACGAACATGT GTACACATAC TCATAAACAC TCACACACAC C
phen32	125 bp	TACACACACA CCACATCATG CATAACACACA CATCAATGCA ATGCATACAC ACATACATAC ACATACTAAC ACATACTACTC ACACACACGC AGAAATTATG CATGCATCAT CGACATTGGC ACGCA
phen33	133 bp	TACACACACT AACACACACA CATGCACACA TACACACAGA CACATGCACA TATACACACA CATAACACACG CATAACACACA CATAACACACA CATATACACA CACATGCACA CTTACACACA CATGCACACA CAC
phen40	151 bp	AGCACTGTGA CAACACAGTG GAGCAGCTTA ACACCACAGT GTAGCACTAT GACATCAGAG TTGAGCACTG TGTCACTACT CAGAGAATA TGACTACTA GTAGAGCACT GTAACATCAC AGTCGAGCAC TGTAACACCA CATATGCGCA T
phwn2	128 bp	GAGAGTAACA TAGGCACAGG TGTGGAGAGT AACACAGGCA CAGGTGTGGA GAGTACACAC AGGCACAGGC GTGGAGAGTA CACACAGGCA CAGGTGTGGA GAGCACACAC AGGCACAGGT GTGGAGAG
phwn3	139 bp	CACTGTGACA ACAATGTGGA ACATTGTGAC ATCACAGTGG CGAACAGTGA CGGAACAGTA AAGGAGTCTG ACAGTACAGT GGAAAACAGT GACGGAACTG TGGAGCACTG TGATTACACC ATGGGACATG TTGCACCAC
phwn4	124 bp	GAGAGTAACA CAGGCACAGG TGTGGAGAGT AACACAGGCA CAGGTGTGGG AGAGTGACAC ACAGGCACAG GTGAGGAGAG TACACACAGG CACAGGTGTG GAGAGCACAC ACAGGTGCGG AGAG
phwn7	132 bp	AGCAGCAGCA GCAACAGTAG TAGAAGCAGC AGCACTAACG ACAGCACAGC AGTAGCAGTA ATAGAAGCAG CAGCAGCAGC AGTAGCAGTA GCAGCAGCAG CAGCAGCAAT TTCAACAACA GCAGCAGCAG CT
phwn10	127 bp	AGACCTTGTC TCAACACACA CACACACACA CACACACACA CACACACGCG CGCACACACA CGCACACACA CATATGCACA CACACACGCA CACGCACACA CACATGCACA CGCGCACGTG CACACAC
phwn13	134 bp	CACACACACA TACTCACACA CCTGTACCAC ACACACACAT GCACATATCT GCACCACACA CAAACACATG CGTGTACACA CACATACTCA CACCTGTACC ACACACACAC ATGCACACAT CCGCACCACA CGCA
phwn14	128 bp	CTTCCTCATG CATGAGCTTG CATGAGCTTG CATATGCTCA CATAACACAC ATGTGAGTCT ACACACAATG AGCACACACA CACACACACA CATCACTAAC CGTCTCGGTC TGGCCATCAT AGTCTGGC
phwn15	133 bp	CACACACACA CACCACATAC ACACACACAC CACATACACA CACACACACT ACATACACAC ACACACATTC ACGCACACAC ACATACACAT ACACACACAC ACCACATACA CACACATACA CACCACATAC ACA

phwn16	140 bp	CACTGTGACA ACAATGTGGA ACATTGTGAC ATCACAGTGG CGAACAGTGA CAAAGCAGCA AAGGAGTCTG ACAGCACAGT GGAAAACAGT GACAGCAGAC TGTGAGCACA GTGATTGCAC CATGGAGCAT ACTACACCAC
phwn17	139 bp	GAGCACCTGT GACACCACAA GGGGGCCTTG TGACTGCACA GAGGGGCACT GTGTCACAAC AGTGGAATGC TGTGACAGTA CAGTGGAGCA GTGTGACAAA ACAGTGGAGT ACTGTGACAC AATAGTAGGG CAATATGAC
phwn18	128 bp	ATGCGCAGAC GCACACACAT GAGCATGCGC AGACGCACAT ACATGAGCGT ACGCAGACGC ACACACATGA GCATGCGCGC GCGACACACA CACACACACA CACACACACA CGAGTGGCAA GGCGGGGG
phwn20	122 bp	ATGCGCTGAC GCACACACAT GAGCACGCGC AGACGCACAT ACATGAGCGT ACGCAGACAC ACACACATGA GCATGCGCGC GCGACAACAC ACACGCTCAC ACACACGAGT GGCAAGGCGG GG
phwn22	131 bp	CAATGTGACA TTACATGTAG CATGGTGAAA TCCCAGTGGA ATACTGTGAC ACCACATTGG AGCACAATGA CACCACTGTG GAGCATGTGA CACCACAGTG GAGCACTGTG AAACCTCAGT GGAGCACTGG T
phwn23	126 bp	GCAGTGCTCA CATAACGCGC ACACATACAG TGCTCCATAT AGTGCACACA TACAGTGAC ACATACGGTG CTCACGTACA GTGCTCACAT ACAGTGTACA CATAACGTAC ACACATACAC TGCACA
phwn24	139 bp	CACTGTGACA ACAATGTGGA ACATTGTGAC ATCACAGTGG CGAACAGCGA CAAACAGTA AAGGAGTCTG ACAGCACAGT GGAAAACAGT GACAGAAGTG TGGAGCACTG TGATTGCACC ATGGAGCATG TTACACCAC
phwn25	134 bp	ATGCGCAGAC GCACACACAT GAGCATGCGC AGGCGCACAT ACATGAGCAT ACGCAGACGC ACACACATGA GCATGCGCGC GCACACACAC ACACACACAC ACACACACAC ACACACACGA GTGGCAAGGC GGGG
phwn28	139 bp	CACTGTGACA ACAAAGTGGA ACATTGTGAC ATCACAGTGG CGAATAGCGA CAAACAGTA ACGGAGTCTG ACAGCACGGT GGAAAACAGT GACAGAAGTG TGGAGCACCG TGGTTGCACC ATGGAGCATG TTACGCCAC
phwn30	131 bp	CAATGTGACA TCACATGTAG CATGGTGAAA TCCCAGTGGA ATACTGTGAC ACCACATTGG AGCACAGTGA CGCCACAGTG GAGCATGTGG CACCACAGTG GAGCACTGTG AAACCACAGT GGAGCACTGG G
phwn32 [†]	138 bp	TGCCAGATGA CACGTGCTAT GCCCAGGTGA CACAAGCTAT GCCCAGATGA TACGTGCTAT GCCCAGGTGA CACATGCTAT GCCCAGGTGA CACATGCTAT GCCCAAGTGA CACATGCCAT GCCAGGTGAC ACAAGCTA

[†] Reverse complement of unlabeled file in sequence dataset of M. Kubista.

SET 3 Phased TATA (96%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phen5	123 bp	AGGTCTATAA GCGTCTATAA ACGTCTATAA ACGTCTATAA GCGTCTATAA ACGCCTATAA GCGCCTATAA ACGCCTATAC GAGCCTATAA ACGCCTATAC ACGGCTATAC ACGTCTATAC ACG
phen25	126 bp	AGGTCTCTAA GCGTCTAAAA ACGCCTATAA ACGTTTATAA ACGTCTATAA ACGCCTACAA ACGCCTATAA ACGCCTATAC AAGCCTATAA ACGCCTGTAC ACGTCTACAC ACGTCTATAC ACGTCT
phen41	126 bp	AGGTCTATAA GCGTCTATAA GCGTCTATGA ACGTCTATAA ACGTCTATAA ACGCCTATAA ACGCCTATAA ACGCCTATAC AAGCCTATAA ACGCCTATAC ACGTCTATGC ACGACTATAC ACGTCT
phen49	126 bp	AGGTCTATAA GCGTCTATAA ACGTCTATAA ACGTTTATAA ACGTCTATAA ACGCCTATAA ACACCTATAA ACGCCTATAC AAGCCTATAA ACGCCTATAC ACGTCTATAC ACGCCTATAC GCGTCT

SET 4 phased TG/CA (96%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phen6	126 bp	ATTTGTAGAA CAGTGTATAT CAATGAGTTA CAATGAAAAA ACATGGAGAA TGATAAATAC CACACTGTAG AACATATTAG ATGAGTGAGT TACGCTGAAA AACACATACG TTGGAAACCG GCATTG
phen16	123 bp	ATTTGTAGAA CAGTGTATAT CAATGAGCTA CAATGAAAAT CATGGAAAAT GATAAAAACC ACACTGTAGA ACATATTAGA TGAGTGAGTT ACACTGAAAA ACACATCCGT TGGAAACCGG CAT
phen27	126 bp	AACAGGATTT GTAGAACAGT GTATATCAAT GAGTTACAAT GAGAAACGTG GAAAATGATA GAAACCACAC TGTAGAACAT ATTATCTGAG TGAGTTACAC AGAAAAACAC ATTCGTTGGA AACGGG
phen48 [†]	118 bp	ATGTAGGACA GTGTATATCA ACGAGTTACA ATGAGAAACA TGGAAAATGA TAAAAACCAC ACTGTAGAAC AGATTAGATG AGTGAGTTAC ACTGAGAAAC ACATTCGTTG GAAACGGC

[†] DNA-binding site for CENP-B

SET 5 No sequence (40%)

Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. & Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* 267, 807-817.

phcn1	144 bp [†]	GGCGGGTAAT TGAGCATCAG TAGGTATAGC TAAGCGGGAT GCGATAGTAG CAATAGCGTC AACAAAGAGGA TAGTGGTGAG CAGAGACGTG CTGCGCGCTA CAATTGGCGT GGCAGTAATA GTCCTGTTTA GTAACCTCGGT CTTG
phcn15	145 bp [†]	CGGGAGTAGA ATGCAAGCGA GGCCGGTACG AGCTAAAGAA TGCCTGATGC AGGGATAAAC GTGAACCGTT ATACAGAAAA ATGCAGATAG AGTGACTAAC GGCGAGCTTA CTGCCACGCG TTGAGAGCCA GCACCGGACA CGATA
phcn34	125 bp [†]	CAGCCCCGGC CTTGAACTAT AGAGCAGCTG AAGATAACTT TGAATTTCTG ACCCCTCTAC CTCTACCTCA CGAATGTTAG GACTACAGGT GAATGCCACC ACACCCCTACT TAACACAGTG ACCAA
phcn35	145 bp [†]	GTGAGCTGTG CGGAAGCAAA AGGAAAGCGA TTAACGGTGT GCTGCATGAT AGTAAGTTAT GCCGCGGACA AGCAGCGTCG AAGCACCGGC GATGGCACAG CTGGCACTAA GGCCGGTGGT ATCAGAAGCT CCTGACAGCT TAGGG
phcn36	146 bp [†]	GGCGAAAGCG GTGAACAGTG GGTCGGAGCC CATTATATTG GTTGAGAGGG CATTGAGACC ACGCCAACAA AACATTGGTT GCATGTGATA GTTTAACCAAG CGAATGACTC GATAAGTCTT AAGATAACGG GCGTATTGGT CGTTTG
phcn38	144 bp [†]	GTGAGCTGTG CGGAAGCAAA AGGAAAGCGA TTAACGGTGT GCTGCATGAT AGTAAGTTAT GCCGCGGACA AGCAGCGTCG AAGCACCGGG ATGGCACAGC TGGCACTAAG GCCGGTGGTA TCAGAAGCTC CTGACAGCTT AGGG
phcn45	144 bp ^{†E}	AGAGGGCGGG TGATATGGAA TCGGGGCCAA TGCGCAGTTT AAGGTTACAG ATTAGGAGTG CGCGCGGAGC TGTAAGCTCG TCGAGCGTCC ATTGGCTAGG TGGAAGTGGA TAGAGGCGTC GAGGGTCCGG ACCAGGTAGC GTTG
phwn1	134 bp [†]	GCGGAGGGTG GGGCAATGGG AGCTGCGCAG CAAATACTCC GTCATCAGGG GTTAGCCGCG TGTAGCAATG GAGCTGGTAG CAGGCACGAA TGACTCGGGT TTGGGTGTAA CCTGAATAGC GGCTGTAGTC GTCG
phwn5	144 bp [*]	GAGGACAAGA GCACCTAACT AACTAAAGC CGGACCGTTG GCGCTCACCC TGTGGTGATC AATCTCCCAC CACGCTTTCC ACCTGACAGC GCAGAGTATC CCAGTCAATA TAGTTCCGCA TCAAGACAGA AACGTCAATG ACCC
phwn6	140 bp [§]	CGGGACAGTG TCGACAAGCA AAGGCGTAGG CCTACCGAGA GGCGTGGGAG CAGTGGTAAT CAGGGCAGCG GCGCGAATAG CGTACGTAGC GGGTGTGAGC TATTACAGGC GTTACGATGC CCTATGTCCG GTGTGCGAGC
phwn9	127 bp [*]	AGGGAGGGTT CTGACCTCAG AACTCAGGAG GTGGATCAGA GCCCCAGACT GCTAGACACA TGCCCTGAAA GAGGAAAGCT TGCCCTGCAGA GAGTGCTCTG ACCACTGGAA CTCAGGAAAG AGCTAGT
phwn19	137 bp [*]	TGTGAACAAC CAATCAACGG TGGCAGTGCA GCATGGTCTA TCAGGTTGTA CAGGCCAGAG CGAGACTAAA ATCAATTCCA CACAAACCCT CTTACCAACG TTAGGACCAT GATCTCTCGC ATCTCCCCC AACCCCG

phwn21	146 bp [*]	ACTCAAAGAA CAAAGATCCT GCTAACCACG GCATTAGGGA ACGGGCGGTA CCGATGCCGT TCTGGTCGAC AGCGCATAGC CCCGGTCCAA CTCCGTGCGG CCTAGAACGT TACGTACCCT AGATGCAGCG GAACTCTTGC GTGTCTG
phwn31	144 bp [§]	CGGGCAGTAT TGCGCACGAA CAAGGTTACT TTTGAAGACC ACATCTGTAC ATCTGGGTAG AGCAGGAAAG CAATTGGTGT GAAGTCATTG GCGTTGCTGC ACAGACCAGC GTCACAGTCG TTAGGGAGCC GGAATGCTAT CTGG
phwn33	124 bp [*]	GGGCTGTAGA ATCTGATGGA GGTGTAGGAT GGATGGACAG TATGACAAAA GGGTACTAGC CTGGGACAGC AGGATTGGTG GAAAGGTTAC AGGCAGGCC AGCAGGCTCG GACGCTGTAT AGAG
phwn34	142 bp [*]	AAGACAAAAT ATGCACGATG TCACATGCAG GACCGCCGAT TGTATTGATA CCATTACGTT ATGCGTGGAC GTCGGCTGTA GTCCTAAGCG CACCCCGACC GAGTTCTGTG TACGAAACCT ACCAGCTCCT TCGACGCGAT GT
phwn35	145 bp [§]	GCGGATAACA ATTTACATA GGAAACAGCT ATGACCATGA TTACGCAAGC GCGCAATTAA CCCTCACTAA AGGGAACAAA AGCTGGAGCT CCACGCGGTG GCGGCCGCTC TAGCCCGGC GGATCCCGAG CTGTTTCCTG TGTA
phwn37	136 bp [*]	TAGACCAGGT GAGCAGGAGG CGGACAGCAG GGAACAGTCT GGAGGGCAGG AAAGAGCTCT GAGGAGCCAT AGCGGGTAAA GCTGAGGATG GGTTTAAGCA AAAGCCAGAC CAAGGACAGG AGGATGTGCA CACTGC

[†] Reverse complement of unlabeled file in sequence dataset of M. Kubista.

[‡] Chain length with extra space removed.

^{*} Sequence included in M. Kubista list of Set5 sequences and in table.

[§] Sequence not included in list of Set5 sequences in M. Kubista dataset (but in table).

(ii) Anti-selection sequences from the mouse genome[†]*TGGA Fragments*

Cao, H., Widlund, H. R., Simonsson, T. & Kubista, M. (1998) TGGA repeats impair nucleosome formation.
J. Mol. Biol. 281, 253-260.

35	105 bp	ATCATGGTGA TGTGATGAT GATGATGATG ATGGTGATGA TGATGATGGT GATGGCGATG ATGATGATGA TGGATGGCGA TGGTGATAAT GATATGGGGA GTAGG
19	123 bp	AGATGGATGG ATGATGGATG GATGATGGAT AGATGGATGA TGGATGGATG GATGATGATG GATGAATAGA TGGATGGATG GATGATGGAT GGATGGACGA TGGATGGATA GATGGATGGA TGG
29	116 bp	TGGATGGATG GATGGATGGG TGGATGGATG AGATGATGGA TGGATGATAA GGTGATGGAT GGATGGGTGG ATGGGTGGAT GGATGGATGA GATGATGGAC GGGTGCATGA TGGATG
75	86 bp	TGGATAGGTG GATGGGTAGA TGGATGAATG AATGAGTAGA TGGATGGATG GATGATGGAT GGATGGATGG ATGGATGGAT GGATGG
44	119 bp	TGGATCGATG GATGGATGGG TGGATGGATC AGATGATGGA TGTGATGATA AGATCGATCG GATGGATGGG TGGATGGCCG GATGGATGGA TGAGATGACG GACGGATGGG TGATGGATG
81	111 bp	TGGATGGATG GATGAATGGG TGGATGGATG AGATGATGGA TGGATGATAA GATGATGGAT GGATGGTGGA TGGGCGGATG GATGGATGAG ATGATGGATG GATGATGGAT G
49	116 bp	ATAGATGGAT GAGTGGATGG ATGGATGGGT GGATGGATGG ACGGGTGGAT GGGTGGATGG GTAGATGGAT GGATGGATGA GTGGATGGAT GGATGGATGG GTGGATGGGT GGATGG
62	115 bp	ATAGATGGAT GAGTGGATGG ATGGGTGGAT GGATAGATGG GTGGATGGGT GGATGGGTGG ATGGATGATG GATGGATGAG TGGATGGATG GATGGATGGG TGGATGGGTG GACGG
57	112 bp	ATAGATGGAT AAGTGGATGG ATGGATGGGT GGATGGATGG ATGGGTGGAT GGGTGGATGG GTGGATGGAT GGATAAGTGG ATGGATGGAC GGATGGGTGG ATGGGTGGAT GG
47	126 bp	TGGATCGATG GGTGGATGCG ACGGGTGGAT CGGATGAGAT GATGGATGGA TGATCAAGAT GATGATGGCA TGGGTGGATC GGGTGGATGG ATGGCATGAG ATGATCGGAT GGATGGATGA TGAATG
77	120 bp	TGGATGGATG GATGGATGGA TGGGTGGATG GATGAGATGA TGGATGGATG ATAAGATGAT GGATGGATGG GTGGATAGGC GGATGGATGG TTGAGATGAT GGAGGGATGG ATGATAGATG
23	117 bp	AAAGGGTGGG TGGAAATGGGA TGGAAATGGGG TGGAAATGGAT GGACGGATGG ATGGGGTGGG TAGAATGGAT GGATGGATGG ACGAATGGAT GTGTGGATGG ATGGATGGAT GGATGGT

80	109 bp	AGGTGGATGG ATGGTGGATA GATGGATGGA TGGACGGATG GATGGGTGGA TAAATGGATG GATGGATGGA CGAATGGATG TGTGGATGGA TGGATGGATG GATGGATGG
86	119 bp	ATAGATGGAT GAGTGGGTGG ATGGATGGGT GGATGGATGG ATGGGTGGAT GGGTGGATGG GCGGATGGAT GATGGATGGA TGAGTGGATG GATGGATGGA TGGGTGGATG GGTGGATGG

BADSECS

Cao, H., Widlund, H. R., Simonsson, T. & Kubista, M. (1998) TGGA repeats impair nucleosome formation.
J. Mol. Biol. 281, 253-260.

24	93 bp	TACCACAATG ACTTGGACAC AAGATACCCC CTCATCATCG CACCTAATTG GTATCAATCA GGGGGCCATC ACCCTCCAGA ACTAAAGTTG CTC
26	81 bp	ATGACAGACG ACGCTGTCAA ATATGATAGC ACCAGCCGAG CTCCCCGATG TATGCCACG TTATCTTGTC CTGCTGAAGT T
28	95 bp	AGGCTTACTC GGACATACCC CCCACTAGTC TAACTTTAAC ATTCGAATGA CAAAAACCAC TGAACCTACC TAAGTGCTG ACGTTCTTGT CGCGG
48	117 bp	TCTAGAGTGT ACAACTATCT ACCCTGTAGG CATCAAGTCT ATTTTCGGTAA TCACTGCAGT TGCATCATTT CGATACGTTG CTCTTGCTTC GCTAGCAACG GACGATCGTA CAAGCAC
51	91 bp	AAGATCGCTA ACTATCCTGG ACTTTGCGAC GAAAATCTAG CACTAAAGGC AGGTCAACTC TTGACCACAC CTCGACGACT GCCCGCAATT A
78	97 bp	ACGAATTACA ATAATACGCC ACATAAGAGG TATGACCCGC CCATCACTTA TCAAGTACAG CACTTGCTG CTATTGACTC TGTTTCTTGA GCAATCC
31	88 bp	ACTACTTCGG TATCACAAC TCGTGAATCG CTACAAGTTA CCGGACATAA TCGTTGGTCA CTGGCAGTCC CCATATCTAC CCAAGCGT
32	86 bp	ACCAAGATCC TGTATCAGT CGTAAAGATT AACCAACTGA CCAATAGGAA CACAGTCAGT GGGGGAATCG ATTGAGTACG GGCCTC
33	85 bp	TTGCGCAAGT CGGCGTACTC GTCGACCATT CAAGCTGTTA CCCTAGACTC TCATACCCTC TTACTACTACC AATACCATAT GCGGA
52	99 bp	ACGGTAAGGC GCACACCCGG CTTCTATCCC TCAGCACCGT CTATACCTTG CCGATATTCA CGGCAACTAC ACATCAACCA CTACGCGCTC TCCTCAGGC
63	89 bp	CCGGCTAGCG TTTGTCCCAC TTTTCACTAC TATGCAGACC CTGATGTACC CGACTCGAGT ATCGTACCCT CACCGGCCGT CTGGAGTTG
69	99 bp	TACTTCGACT ACACACGCCA ATTCCGCCAA CCTCGTATCA ATTAACCTCA TGACGTGTAT ACCACCCGAC ACGCAAGGCA TCTCGCAGGC TCACCCTCG
73	95 bp	AACGTTTCCC CGAAAGTAGG CCTGATGAGA CGACACCCTA TCTCATGTGG ACGACTCATA CGCCCCTGGC ACCATTTCTT CATGTAGTCA ACAGG
76	92 bp	TTAGTCCGAA CCAGCAGACG AGTACCCTAG CTGTCGGAGG TATCTATCAC GTGCCCCAGC TGCCACTCAT AATACTACAC GAGATCCGGC CC
84	102 bp	AGAGGTGACA TGCGACTGAG GACATCTTAT CAGTAGTCAC AACAGCAGAT GGTGCTGCAA TGCCCAACGA TTACCGTCTG CCGCATAGCA CAAGACCCGA AA
85	99 bp	AACACGTCCC ATGTCACCCC TGCCCTGTTGG ACTCTCAAAA TGCTCACTAC GCTGAGTATA TTTTATCACA CATGTGTGGT GGACCCCGA ATGCTAACT

27	87 bp	TACAAAATTA CGCAAGGCTA TAATGGCCCG AAGGCACGGA AGACCCGGCA AGCAAACGCTA ACTGGCGCGC CAATCCTATG CTTTNNNT
30	88 bp	AACCAAGTGA CAACTAACAC ACACAGGTGA GTTATCAGCG GGCCGGCATC CAGTTTAGGA GTGCCACGCC CATCTATGTT AGATACAA
43	108 bp	CACCACATAG TTTGGTGAAG CCGGCTTCCA GCAAGCACAG TCGACTAACA AAAACGAAAT ATCTGCCCCG GGAGACGAGC TCCTTGATT ACCGCTCTTG CCAACATG
53	87 bp	ATCACACAGG AAACAGCTAT GACATGGCAC ACAGAAAACA GCTATGACTG GGAAAACCTT GCCGTCGTGA CCGGGAAAAC CCTGGCG
55	89 bp	TTATACCCAA AAAACCAGGT GAACAGAACG GCCATGCAAT AGTTCCTGCA ACAAACCTGC CTGAAATACG TACACGGAAT CTAAAGTAC
56	77 bp	TTCACACAGG AAACAGCTAT GACTGAGAAA ACCCTGGCGT CACACAGGAA ACAGCTATGA CTGGGAAAAC CCTGGCG
59	96 bp	TAGATTCAAG TGACGGACGC GTCTGCCGAC GCCCACTAGA CTGTGCGAAT GGTAAAAATG TTACTGAATA TCCTGGAAAG AGCACCTAGC TGCAAG
66	101 bp	TGTGACGGTT ATGCAACTAG CACGTGATTT TCGTAAGAGG TCACAAGATA ATACGGCGAG ATAACTTAAA CCATCGGACC CGATGACTAA CGGATACACT A
67	84 bp	CCAGTAGGGT TTGTTCATAC TGAGTAAAAG TCTGTCCATA ACTACCTACA TGGAACGCAA TCAGTCACAG AAATTAGCGC CCGA
74	87 bp	AAGAAAACGA TAGAGATCAA ACAGCAAACC TAGATAGTAA TACGACGGAC TCAAACCTCCG GTCGGGCGAG CTGAGGTTGC ACCTAGA

† Names are denoted pasHC# as the plasmid counterparts.

(iii) Synthetic SELEX sequences*

Lowary, P.T. & Widom, J., New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. (1998) *J. Mol. Biol.* 276, 19-42.

10nvp	233 bp	GCGAGCTCTA GATACGCTTA AAGTACTTAT CGGCGGCGTA CAAGCGTTTT AGGCTCCAAT TGCCTAGACC GCGGTGCCGG TAAGCTAGCG GGTCGAATGA TTGTACTTCC CCCCTTCTGT CTTAAGCCCC CAGCTCATCC CTCTGTATTA TTCACCCAAG GTCTCGCGCA ATTCCCCTGA GATTGGGTAC TAATGAATCC GACCCGTGCG TGGAAGCCGA GAGGCGATTG CGT
11nvp	229 bp	ATGCACTAGA GGGTGACGAG CGCGATGCTG GGCCGACACG CTGGATGTAC TGTTTTTCATT AGTTGTAAAG GTGCCCGCTG GAGGCAGCAT CACAAGTTAT AATTCGTCCA CTTTAAAGAAT GTCGCCTTCG GACTACCAGG GCGGAACCCG ACCAAGGGAC CGCTTATCTC GCGTTTAAACG GCGTTTGCAC GCATCTAGAG TACGCTACGA CATA CGCCGA ACGTAAGCC
13nvp	233 bp	TACATAGATT GTTGTACATC TCTAGACCCG CTTAGCGTGC GCTACTGTGC ATTCTGAGCG GTTCACAACA CGGGTGCCCTG GGTTCCGGCC CTCAACTGCG AGCGAGCCCC TGGTGATCCC TTGTTTGGCC CTGCTCGTAC TGCTTCGGCT CCATCCAAGG GGTTGAGAAC GTGGCATGCG CGAAAACGAG CCATATTTAT TTCAAAGCGA TCCTCATTCG CGTTCCTGAC TAT
16nvp	233 bp	GATGCGTATT GCATACAAGG CAAGATAAAC GTCGCAGAGC TACCATATCG TGCACAGGCT GCGCAAAGTC GGGTGCCCGC GACGACGTGT CTGGTTTCTC AGCTATACTG GGACTAGCCC GAGAGTACTA GAAAACCCGT TTGGAGCCTC ACGACCAAGG ACAAGAGGTA AGTGCGTAGA TTGCTGTCTA AGCCAGTTTA AGCGTACTTA GAAGGCCCTA ACCGACGGCG GAG
18nvp	232 bp	TTACAATTTA CAGTGACTTG GACACTTACT GCTACAGAAC CACAGGCAGC ATGCTTAGCC TGGCGATTGC TTGGTGCCAA TACAACGCAC CAGGCGCTGT AATCGACGTA ACTCAACGAG CCTGCCACGC TAAGCATACT TAGAACGTGC CTAACCAAGG GGTGGAGCCT ATCGGGCTGT GTGATTAATAA AATGAACTCT CTCGCTCTGG AGTCTTCCCG TACCCTTCGC TC
22nvp	231 bp	CAAGGGGGCG GGATGGGCTC TACAGCGAAC TATCACCCAT TCCGCGAGCG GCACAGTTCC ATCCACTGCT GGGTGCCGGA GTGTTGCGCT TCATACGGTA ATTGCCGGGG TACTACCTGC CGAATGTGGC TTAAATGACA TGACGCCGT GTTCCAAGGG TCTAGAGCAT CGTAGCGCCT TCCCCTGAG GAAAGTAGAC TCGAACACGG TGCATCCCTA ACCACGGCTA A
24nvp	228 bp	GACTTCGGAA CGTAAAATTT CTGCCCTGG CTTCTGCGGA CCGTCATCCC CCGTATGGAG AGCCCCGCA CATGGTGCCA ATAGACATGT CTTTCTACAA ATTCAAAACG GGACTACAGT TGAGCTAGGC TGACGGGGAT CTATGAAGCC CCAAGGAGTG CAGGGGTGTA CAGCTCACCG CGTACGTCAC TTAAAGCGGG TCTAGGCGTC ACTAAGGGGC CGGTTGTA
25nvp	232 bp	GATCTATAGA GCTCTAGAAT ATTTATGGAG CTGAAAACCTG CTCCCCACA CCCTATACTG CCCGCGTAAG AAGGTGCCGA CTATGTCTGC GGAAATTAAG AGGTATTAATA CGTGCAGCTG GAATGCTCTA GACCCCTTTA GGAGCCGCTA GAGGCCAAGG TTAGCATATT AGCTCGGTAG TTtGGCCTCA GTCAATTGGG TTCACTCCGA ACTGGACATA CCGATTGGA AC

26nvp	229 bp	TGGGCAAAC T ATCAGTCACA AATCACTTGC TTCGCCCAGA GCACGCCCCG CTTTCATTAA ACTGCCCGGT GGTGCCTCTA GTGTACGCCA GACGGACGTA GAGCCTATTC TGGAGCGTTT CCACATAATC CACGAGACAG AAGATACTCA TACCAAGGGA AATCTAGTGC AAACGCAAGT GACGTAACAG TATAACCGCG CGGTGGCTAA CGACACCAA ACTCCGATG
28nvp	226 bp	GAAGAGAATC CCTTCTCTC CCAACTAAAT GCTTAGTTGG CGTCGGTGCT ACCCCTCAG GAAGGGTCTT AGGTGCCTCT AGACCCCTGC AGACGGCCGC TTTCACTGAG TTACTACTGGA TTGACGTTGT CCCACCCGGC GTTACTATCC AAGGGTCTGC CATTACGCAT TGTAGGATGT TTCAAGCTTT GACTGGGTAT TAGGAGGAAT CTCACGGTCA AACTAA
29nvp	231 bp	TGGGAGGGAC CAGTCCGCGA TGGAAATTAT TAACATGAGC ACCACAATTT TCCGTCCAAC ACAGGCACTC CTGGTGCCTG AATGAGGGGC TTAAGTACC CTAGTTAGCT GTGTAGCGAA CTCTAGTGTG TCCTAGAGTG CCTCTCGTGC GATTCCAAGG GGTTTTGGCC TGGGGGATAG GCCACATAGC GGTACTATAG GAGTCTGAGC GGGCGAGTTT GCCTCTAAGA C
2nvp	232 bp	GAGTCAGCCA GCTAGAACGC TCTAGACACT GTTAGCGTGA CGTAGGCCAT TAGATAAGCC CCTACGGAAC GCGGTGCCGG CGTTGTAACG AGGTTGTGTT TTACACGGGA ACATTATGTC GCGTAGCGGC AATGGCGTAA CCCACCTTAC TGTTCCAAGG TCTCACCTCC GAGTGAATCA TCCTCCAGGG TGTGCTCCTC CGGGGTACCC CGCGATGTCC GTGGGTCCGGC GA
30nvp	229 bp	TGTGCGTTGT GTAGTCACGG CGTTTGTGTC TCGATGAAAT CGGAAACCCA GAACTTAAGC TGTGCTCCAC ACGGTGCCCC AAAAACGTTA GTAGGCCGGT CTACAGCGTG CTAAAGCACT CTAGAGCCGA TCGTGACCCT GCTATGCTCT GCCAAGGGCA CCTGGTGTAC CCCTTGCACG CAGGTCTGCC TCCGTGTATA CAGGACATTA CCGTGCGCTG TACTATGAT
34nvp	231 bp	rTCgACGGgw gGGCCGCTg cATGcCGTTA CCTcATCTTg CtCCGCTCTA ACCAC TTCAT TATCGAGCAC AGGTGCCCAG ACGATTTAAT TTcCGTgGGA CGGCATcCTG CGGAaGACG TATTTGTCG TAGCAATGGC TGTTAAAGGT TCTCCAAGGA GCCTTCTAGA CCGGCTTAGA GTCATCAGAA TcCAATTCTA AGcGCACCGc raATACTACT GAGGCTAggg G
37nvp [#]	229 bp	ACGAATGCAC gGGgcCCGGT CTAATGGAGG CGTAATGGAA GGCAGCTAgG AGAGaAATGC TGCATTCAGT GTCCTTGGca tTtAAGCtA TACTAGACGg CCCCGGCGGg GACTGGAACA CTCGGATAGG CATCAGGAgT AAGCATTGCA ATAGGCACCC AGATTATTGc CCAGATTGTG CTaACACccg gagGaGttAc GacAGTgcaA AGcacATCgG ATCTAGcCA
38nvp	230 bp	cCgGGAAGCs gcgCCAgGat TTCCTccctT ATtgGCTCTA gGTATGcTta AAcgGCTtgG cgcGTTGtTT GCGGTGCCCC TGTGCATGCT AGAGCctTTG GTTGAATGTG CATCAgCCTA GTATTGCAGG AGCgcccCGa GCCAACTGTA CAcCCAAGGT ATTCGTGTGA ACTTCgGcAC tTTTGTgGcTG TcTtGAGttc gCGTTAgCgg tTTTTAGGgg tAAAtgACCA
40nvp [#]	231 bp	GcGGAaGGT GGtggGGGTT GTTCAGGcTt CTAGTAGTGC TTAAAGTGCG TTAGAGGCC AgcGCAGcgc tCCCTTGGCA CgGGtcccC cACTGAATAG TaCCTATccT ATTTTcGccG TtttacCGAG GGgtTTaGGA GGcCATACAA AAtGGCACCG TTAAGAtcG ccaggAGGcA CGcCGTgGat cGaGatCTgg CtgcgAcAta aTgcCATgAt TCctgAcCaT t

41nvp [#]	230 bp	AGCTGGAAAG TGTAATAGGG CGGGCGCGGG GCGGACACTA TTCgAAggCT gTGTGcATCA TAGTCTGGCA CCTTGGCAGA ATGAGGAGAT ATTGCgCCCT GTACCGGAAA ACgGCTCAGA AcGACGTAAC AGamgcccTA AGcgCACCTA gGGGCACCTA tCTCGTTCGA ACGGGTGTcC ACGCgGTTTC TGcgAATCCT gAACcGGGGA TAGCAGTAGG TgACgGCACT
52nvp	232 bp	GGACCCGtTA TtgGAATTGA GTCGTATGTC CGACgGCGCG TTGAAGTGGC CGAtTTTTCA ACTACACgGC ATGGTGCCCA CCTCCATAGT CGTGATCTGA ATTcCGCGGC TACATGCGcT GTGTCTTCTA GACCCTCTAG GAGCACGTTA GAgGCCAAGG CAAGCACGGt cCCGGTAACG AACgGGGTTG CACtGGTACG CacCCCTAGA TgGGcAGCTT GGAACGcCTA CT
54nvp	230 bp	AGAACCTTCT ACATACGCTT AGGGTGAcGc ACTCCTcAtT TCTGAGcCTC GtGcTtTAgg gCGTATAAgG AAGGTGCCAG TCTCGTAGTG TTATAAGCGc GCCCCACGTT GTCAAACGGG CGGCGATCTC TCTTGGCCGT GGTGTTTTATT TGCCCAAGGT TGCGTGCGTA CCaTGAAGTC TTTTTATTGA CTCACCCAAC GAACACACTA GTGCGTACCT cCCGCTTACg
5nvp	231 bp	GGCTAACTAG GCATCATTAG GTTGTGTGTT TTAGGCCTCT CCCAGCGTCT AACAGTGCTT AGCCACACCG CTGGTGCCAG TGCAGCACTT TACGATACAA TGCGGATGCT TATGGCGAAT TCCCTACTGG AGATCCCTCA CTACTTGCAG TTGCCCAAGG CGTACTGAGG GTATCGACAG CACCACGTAA AATTTTCCAG TGGGGACTCA GGAGTCTTTG CGAAAGACTA G
602nvp_rev_rev [#]	232 bp	GGCAAGGTCG CTGTTCAATA CATGCACAGG GTGTaTGTAT CCGACaCGtG CCTGGAGACT AgGAAGTAAT cCCCTTGGCG GTTaAAAATGC gGGGGACAGC gCGTACGTgc GTTTAAGCGG TGCTAGAGCT gtCTAcGACC AATTGAGCGG CCTCGGCACC gGGATTCTcC AGGACGGcCG CGTATAGGGT cCATCACATA agGGATGAAC TCGGTGTGAA GAATCATGCT TT
603nvp	234 bp	CGagacATAC ACGaATatgG CgTTTTCTTA GTACAAaTCA CCCCAGCGTG AcgcGTAAAA TAATCGACAC TCTCGGGTGC CCAGTTCGCG CGCCCaCTA CCGTGTGAAG TCGTCACTCG GGCTTcTAAg TAcgcTTAGg cCACggTAGA gggcaaTCCA AGGCTAACCA CCGTGCATCG ATGTTGaAAG AGGcCCTccG TCctTATTAC TTCaAGTcCC TGGGgTACCG TTtc
604nvp	233 bp	GTCGTCTGCA CAAATACGGA ACCTTTGGty CGgTTTTGCT CTAAGACACG GGCgGCACTc TAACTAGGCT TAAGGTGCCG TAAAGCCGAT TCTGAGTATA AtcCTGTCTC TTGACTTgTc TCGGGCTAAA GTTGACTcGA TGtCGTGACA TCACCCCAAG GCGATCTTGT TTGTAGTcCC CCGTsTATCG TGGGGCCGCG AGACATAGCA kCCAATAGGC ATGATGAATA GTC
605nvp	231 bp	TACTGGTTGG TGTGACAGAT GCTCTAGATG GCGATasmGA CAGGTCAAGG TTCCGACgac GCGGGATATg GGGTGCCTAT cGCACATTGA GTgcgaGACC GGTCTAGATA CGCTTAAACG AcGTTACAAC CCTAGcCCCC TCGTTTTAGC CGcCCAAGGG TATTCAAGcT CGACGCTAAT CACCTATTGA GCCGGTATCC ACCgTcACGA CCATATTAAT AgGACACgCC G
607nvp_rev_rev	227 bp	CTTTGGCGGA GAATGCGTAG GTGAGTGttt tgGAgACGgg aAgGCATaaA ACACCcgGAG CCTAACGaGC GCCTTGGtAt cCGGaacggg tTCTAGTACC cGTTAGCgTG gTTTAGAGGG gCAAAGGAAC ATCTttcccc CCCCAGATA CgGGCACCGt TCgGAcCCTg GTTAGTcCAG TGCTACTGCC GGTTCCTAGC CTAAGGcGAG GATGcCTATT AGaAAGG

609nvp	230 bp	TAATCTTAGC CGGTGTGCAT TACAAGCACT ACAGTGACGG ATGATCTCTA TGCGGATCTT TcAATCATGC GGGTGCCGAG TAgcGTgCcG caGTTGATGG TTCAGCCGAC GAAGCAGGCg GAAcCCACTC CTAACGATGC CCACTACTAT cCTCCAAGGG GCTCTATAcC CCTTAAAgT Gc+tGCGCTGG GCATTGGCTC CgTTTTTCCCC TGTCTTCCAC CGTGCAGTAA
611nvp ^{#†}	232 bp	GaACGCaAAT CTCAGATgGT gAGTAgGGGG CAGTTGAgGG GCTTGGTCTG TGTTTAAACGC GCTTcTTCAG GGCCTTGGAG CTCTCTCTTG AGCTTGCTGA GCAAtcgTCG ACATTAGCtT TCTGCGTGAg CGGCGTCATT AGTGAGAATG CCAAGGCACC GGCATTGCGC gTGAgATGGG CTAGGCaaCG cgCTAAGcgt ATCTaGAACg CGCTACGAAA CATACGCAtG TT
613nvp	232 bp	TAACTAGCAT GGACCCGATT GTACGCCGTA ACACTCGCGC TCGGTGTACC TTAGTcCTAG TATAGGCGTA TTGGTGCCAA TTGCGCcGAG CgTTGtaGCC GtGGAgTgTc GttGtCGtAT GGGGGCTCTA GATGTGCTTA AACGGGTGGG TsGcCCAAGG CcTGCATTAT TGC GGAGTAT GGCCACGcTT CCGTGCCGCT CCcTAGTcG ACCcTtcACA CTCTCACTCG CG
615fnvp	229 bp	ACACTTGTAAC AACAGGACGC CGTTATCTTC CCCTCCAGTC GGGACCGGCT CTCCAACCAt GCCAAGCGAG CGGTGCCAGC AAGGATTCGG TCTCTACGTG CATCTAAGAC CCCTTAGAGT GCTCTAcTCG CTCTcGCTAG CTcCGCGAGA ACCAAGGCCG GATTaATTGG CACGGGAGGT TTCTACaaCC GCACTACGTG TGCTGACGTT CGAgCCCCCT CAAACATGt
618nvp_rev_rev	232 bp	CgGACTAAAC GAACCGGtTc GtTTACAtTG TGCgaAgGGG AtTGGCGTTC TACGACAAAt tCtgGGAACA GCCCTTGGTG GCTCTAGGgA GCTCTAGATG AGCTTAGCgG TCTGTAGACg CAATTCTGGG cATTTCCGCT CGTCTGGatC tGTAGGCACC tGCTCGGTCA gGTACaActT TgaAAAAGGCG cCAGTTATGA GCcgGTATCG TgGGCTTTgG GTAgcCCAcc CC
620nvp_rev_rev [‡]	231 bp	GggACCcAGC GTGCCcAGCG gTCTCTAGaT AcgcttaAAG TACTCTACGC ACTAAAaGAG GcAGCTTGgc CGGGTGCCGC GcAGTGTAaA GTGgATACgc tGGGccTGAC GTATCCGAGG ACgCTCGCgT GTCACTcTTT GCcTTGCCCC CCCCCAAGGA GAAGCCTGAT ACAAacCCAT TCTGCCCGCG AGTCAGTCCG GTCAGaGTGA TACGGCGACC GgtAGATCTT C
621nvp	232 bp	TACAGGGCGG TTCCCTAGCT GTTGcATAAA GGCCGGTGGT TCGTCGCCCC CCCTTAAAGA TAATCCTGCG TAGGTGCCCG TGGAGCCGTT CTATTGACCC TTTACGATCG TTTAAAGGGG TCTAGAGCTA GCTAGCTTGT TcAGTCTAGC CCCACCAAGG ACTgGTgGCG TATCAGTTCA CGgcCTCCcg GGTgGTCgGG TACtTTagGA ATATCCcgtT TACATCCGAC AT
624bnvp [#]	231 bp	ACTCTAGCAA CTGGTgCGgC GACTcGGgaa CGATTAGGTT GCAATCCTAG ACACCCATCC AGAtTGCAct GTCCTTGGac CCAGAGCCCg CgGAgGgACC GATTTCTAGT TTcCTAGTCT CTGGGTAAcC GAAGcCCCAT ACGGAAACGA GCCAGGCACC GTGCAAAAACG TcCAGAGTGA gCTGTcCTAT CGTTTAAAGCG GTCCTAGAGC CTCTAAGCT TGCAACCACA A
626nvp	230 bp	xxxxxGGwGG GAATTTcGwG ATwGTTAwGC TCCACCAwGC CTTGTcGCTC TgCtTCGATG AGAGTCGACC GGTGCCgGGT AtTGtATCGC CGCTTGACTC TGGAATgGCG TCTACGTAGC gCTAAGCACT ATTAGAGTCC TCTCCtGCAT CTCCAAGGAT AGtGCCTATA AATCGTCCAC CTGTgGCATT ATcCGGTGCC CTTTCgGgAC gTGCTTTTTA GTGTCgGGTG

628fnvp_rev_rev	232 bp	TGCGGTTGGG CTAATCGGGG GCTTGGCGCG TAAACCTACC GTTCCGACAC CTGTTCAAAG TTTCTGACCA GTGGTGCCCTG GTCTCTGAAC TGCTGTAGAG AGGCC TAGAC gTGCTTAGCG TG CAGTAGAA CCGATTCTAA ACCGTTTTGA GACGCCAAGG GTTAATGCGC ATAGgcccTT GCATGGCCCC TTGcTCAAGC CGGCTGCCAA CCAGGGccAG AGTCTCCGTT TG
69nvp	231 bp	TCgCAtCcac CCTGTgAAGT gGAGGGTCGA CGTTAAatTTA CCATaccCCG cATAACACGT cACGGCTAAC AGGTGCCAAT CTCCTCGGAA AGAATATcAG GatCGcgtTA cCGCACgGGG GGAGCTCTAG AGTGGcTTAA ACGACCGCAG acGCCAAGGG GAACAGGCTG cAGGAGGTgG CCGgGACtCt tcctgCgATC GAACGCTcTc ttGTAGAATT GGATcCTCgG A
6nvp	232 bp	TGGACACTTT TATGTGTTTA CTCGACCTTC CTCAAACctc GCCTTGGGCT GACTCGAATT TTCCAGGCTG CAGGTGCCCGG TGGAGCCTCT AGGTGTGCTC AGAGGACGGT AGAGCCCGAA TCTAGTGCTC TCCCGGTAT TATGTGCCT TCCCCAAGG GACGGCGCAA GTTCCACACG AGCAACGCAG ATCTGGGTCC TGATCACTCG AAAGCCGTCG CTTC CCATAG AC
73nvp	230 bp	GAGAGAGCTG ATAGGCacGC ggacTCACgt GGGTCCCTct TACGGTGGTG GCAGGAGGGC TAGGATGATT GTAGGTGCCT CTAAGTGTGC TTAAAACgaT cCACGGACcG CTcCTGTcGg TTTCTACACG tGTCGGACTT AATCAAAGCT CACTACCAAG GGCgGCgACG TAGccgaTAT CCGCaAAACT ACCATgaGTG GgCAgCctCc GcGcGaCtcc gGGATCTgGA
7nvp	233 bp	GCACCTGCGT CCCAGCCTTC TCCAGATGAG TTTCCACAGT GTCCCTCAGT TTCACGCTTA AAATGCTGCT TTCGGTGCCC CTGCGTGTTT AGGCACTACT AGAGCATTCC ACGGCTTTCT TTTACACTTG ATGGTCTACG CTCTTACGTT CCATGCCAAG GACTGAGGCC TAGCACGATC TAGCcTcTCC CGTGCTGCCG GGAAAATACG CGTCCATCGT TGCTCTTTCC TCT
83nvp ^{#§}	228 bp	ATGCGGATAC AGCAACGGAT GGATCGGgAC AGCCAATCGG TGCCCTCCCGT GGGAGTAAAT CGACAGGCAT ACCCTTGGAG CCCGCGTCTA CCCGGTGCTC TGAAACggcC CCAAATGCC TGTAAgGCAT CCTAAGCATA CTCCTGCGAT aCTAGGCACC TATGGAAGAT TGGAGTGGGT GTCGGACAGT TCATTTAGCG CAacGGGAGA CCCGgcCGct waGCCCTA

* Single and double underlined sequences are restriction fragment sequences used to ligate segments extracted in selection experiments.

Reverse complement of restriction fragments.

† Restriction fragment sequences overlap with reverse complements.

§ Chain contains second restriction fragment.

‡ Sequence of reverse complement is identical to original sequence.

(iv) High-resolution positioning sequences[†]*Crystallographic DNA sequence*

Davey, C.A., Sargent, D.F., Luger, K., Mäder, A.W. & Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319, 1087-1113.

pd0285 146 bp ATCAATATCC ACCTGCAGAT TCTACCAAAA GTGTATTTGG AAAGTGCCTC
ATCAAAAGGC ATGTTTCAGCT GAATTCAGCT GAACATGCCT TTTGATGGAG
CAGTTTCCAA ATACACTTTT GGTAGAATCT GCAGGTGGAT ATTGAT

pd0286 146 bp ATCTCCAAAT ATCCCTTGCG GATCGTAGAA AAAGTGTGTC AAAGTGCCTC
ATCAAAAGGA AACTTCAACT GAATTCAGTT GAAGTTTCCC TTTGATAGCG
CAGTTTGACA CACTTTTCT ACGATCCGCA AGGGATATTT GGAGAT

pd0287 147 bp ATCAATATCC ACCTGCAGAT ACTACCAAAA GTGTATTTGG AAAGTGCCTC
ATCAAAAGGC ATGTTTCAGCT GGAATCCAGC TGAACATGCC TTTTGTAGGA
GCAGTTTCCA AATACACTTT TGGTAGTATC TGCAGGTGGA TATTGAT

Bao, Y., White, C.L. & Luger, K. (2006) Nucleosome core particles containing a poly(dA·dT) sequence element exhibit a locally distorted DNA structure. *J. Mol. Biol.* **361**, 617-624.

pd0755 147 bp ATCAATATCC ACCTGCACAT TCTACCAAAA GTGTCAAAA AAAAAAAAAA
ATCATGATAA GCTAATTTGG CTGACTCAGC TGAACATGCC TTTTGTAGGA
GCAGTTTCCA AATACACTTT TGGTAGTATC TGCAGGTGGA TATTGAT

TG-pentamer

Shrader, T. E. & Crothers, D. M. (1989) Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci., USA* 86, 7418-7422.

TG-pentamer 190 bp TCGGTGTTAG AGCCTGTAAC TCGGTGTTAG AGCCTGTAAC TCGGTGTTAG
AGCCTGTAAC TCGGTGTTAG AGCCTGAACT CGGTGTAGAG GCCTGAACTC
GGTGTAGAG CCTGAACTCG GTGTAGAGC CTGAACTCGG TGTAGAGCC
TGAACCTCGG GTTAGAGCCT GAACTCGGTG TTAGAGCCTG

pGUB

An, W., Leuba, S.H., van Holde, K. & Zlatanova, J. (1998) Linker histone protects linker DNA on only one side of the core particle and in a sequence-dependent manner. *Proc. Natl. Acad. Sci., USA* **95**, 3396-3401.

Kassabov, S.R., N.M. Henry, M. Zofall, T. Tsukiyama & Bartholomew, B. (2002) High-resolution mapping of changes in histone-DNA contacts of nucleosomes remodeled by ISW2. *Mol. Cell. Biol.* **22**, 7524-7534.

pGUB 183 bp GATCCTCTAG ACGGAGGACA GTCCTCCGGT TACCTTCGAA CCACGTGGCC
GTCTAGATGC TGAATCATTG TCGACACGCG TAGATCTGCT AGCATCGATC
CATGGACTAG TCTCGAGTTT AAAGATATCC AGCTGCCCGG GAGGCCTTCG
CGAAATATTG GTACCCCATG GAATCGAGGG ATC

Table S3. Properties of DNA base-pair steps used in knowledge-based potentials**(i) Number of base-pair steps included in derived ‘energy’ functions[†]**

Step	Olson <i>et al.</i> [*]	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	88	160	160	118	160	118
CA	110	177	166	143	166	143
TA	134	236	232	212	238	218
AG	106	168	157	133	157	133
GG	97	146	139	95	139	95
AA	129	169	169	166	181	178
GA	117	175	173	146	173	146
AT	140	236	230	202	234	202
AC	137	189	188	168	188	168
GC	86	182	164	140	164	140
Generic	1840	2862	2770	2374	2798	2402

[†]Counts exclude terminal dimer units and steps with single-stranded nicks and mismatches within the selected sets of high-resolution structures.

(ii) Average values and dispersion of Tilt, $\langle \theta_1 \rangle \pm (\sigma_{\theta_1})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> [*]	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	0.0 _(4.2)	0.0 _(3.9)	0.0 _(3.9)	0.0 _(3.2)	0.0 _(3.9)	0.0 _(3.2)
CA	0.5 _(3.7)	0.1 _(3.1)	0.2 _(3.2)	-0.3 _(3.2)	0.2 _(3.2)	-0.3 _(3.2)
TA	0.0 _(2.7)	0.0 _(2.7)	0.0 _(2.7)	0.0 _(2.8)	0.0 _(2.7)	0.0 _(2.8)
AG	-1.7 _(3.3)	-1.1 _(2.5)	-1.0 _(2.5)	-1.0 _(2.5)	-1.0 _(2.5)	-1.0 _(2.5)
GG	-0.1 _(3.7)	0.7 _(3.6)	0.6 _(3.5)	0.4 _(3.6)	0.6 _(3.5)	0.4 _(3.6)
AA	-1.4 _(3.3)	-1.1 _(2.5)	-1.1 _(2.5)	-1.1 _(2.5)	-0.9 _(2.6)	-0.9 _(2.6)
GA	-1.5 _(3.8)	-1.5 _(2.8)	-1.6 _(2.8)	-1.4 _(2.8)	-1.6 _(2.8)	-1.4 _(2.8)
AT	0.0 _(2.5)	0.0 _(2.4)	0.0 _(2.5)	0.0 _(2.3)	0.0 _(2.4)	0.0 _(2.3)
AC	-0.1 _(3.1)	0.6 _(2.9)	0.6 _(2.9)	0.7 _(2.9)	0.6 _(2.9)	0.7 _(2.9)
GC	0.0 _(3.9)	0.0 _(3.6)	0.0 _(3.8)	0.0 _(3.9)	0.0 _(3.8)	0.0 _(3.9)
Generic	0.0 _(3.6)	0.0 _(3.1)	0.0 _(3.1)	0.0 _(3.1)	0.0 _(3.1)	0.0 _(3.1)

(iii) Average values and dispersion of Roll, $\langle \theta_2 \rangle \pm (\sigma_{\theta_2})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	5.4 _(5.2)	5.5 _(5.7)	5.5 _(5.7)	6.0 _(5.8)	5.5 _(5.7)	6.0 _(5.8)
CA	4.7 _(5.1)	5.1 _(5.0)	5.2 _(4.9)	4.9 _(4.9)	5.2 _(4.9)	4.9 _(4.9)
TA	3.3 _(6.6)	2.5 _(5.8)	2.4 _(5.8)	2.6 _(5.7)	2.7 _(6.1)	3.0 _(6.0)
AG	4.5 _(3.4)	4.1 _(3.7)	4.2 _(3.5)	4.2 _(3.4)	4.2 _(3.5)	4.2 _(3.4)
GG	3.6 _(4.5)	5.0 _(4.6)	4.8 _(4.6)	5.0 _(4.7)	4.8 _(4.6)	5.0 _(4.7)
AA	0.7 _(5.4)	0.7 _(4.5)	0.7 _(4.5)	0.7 _(4.5)	1.7 _(5.7)	1.7 _(5.7)
GA	1.9 _(5.3)	1.9 _(5.4)	1.8 _(5.4)	2.6 _(3.9)	1.8 _(5.4)	2.6 _(3.9)
AT	1.1 _(4.9)	1.0 _(3.7)	1.0 _(3.6)	0.9 _(3.7)	1.0 _(3.6)	0.9 _(3.7)
AC	0.7 _(3.9)	1.6 _(3.3)	1.5 _(3.3)	1.5 _(3.4)	1.5 _(3.3)	1.5 _(3.4)
GC	0.3 _(4.6)	1.2 _(4.6)	0.4 _(4.0)	0.7 _(3.7)	0.4 _(4.0)	0.7 _(3.7)
Generic	2.7 _(5.2)	2.9 _(4.9)	2.9 _(4.9)	3.0 _(4.7)	3.0 _(5.0)	3.1 _(4.8)

(iv) Average values and dispersion of Twist, $\langle \theta_3 \rangle \pm (\sigma_{\theta_3})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	36.1 _(5.5)	34.4 _(3.7)	34.4 _(3.7)	33.9 _(3.7)	34.4 _(3.7)	33.9 _(3.7)
CA	37.3 _(6.5)	35.0 _(4.9)	35.5 _(4.6)	35.8 _(4.6)	35.5 _(4.6)	35.8 _(4.6)
TA	37.8 _(5.5)	37.4 _(7.3)	37.5 _(7.3)	37.6 _(7.5)	37.1 _(7.7)	37.1 _(7.9)
AG	31.9 _(4.5)	32.5 _(4.6)	32.6 _(4.6)	32.3 _(4.5)	32.6 _(4.6)	32.3 _(4.5)
GG	32.9 _(5.2)	33.3 _(4.5)	33.4 _(4.5)	32.7 _(3.9)	33.4 _(4.5)	32.7 _(3.9)
AA	35.1 _(3.9)	35.1 _(3.9)	35.1 _(3.9)	35.1 _(3.9)	34.3 _(4.8)	34.3 _(4.8)
GA	36.3 _(4.4)	35.5 _(4.2)	35.7 _(4.1)	35.5 _(4.2)	35.7 _(4.1)	35.5 _(4.2)
AT	29.3 _(4.5)	29.8 _(4.0)	29.8 _(4.0)	29.8 _(4.2)	29.5 _(4.4)	29.8 _(4.2)
AC	31.5 _(4.2)	31.7 _(3.7)	31.7 _(3.7)	31.5 _(3.6)	31.7 _(3.7)	31.5 _(3.6)
GC	33.6 _(4.7)	33.7 _(5.0)	34.3 _(4.6)	33.5 _(4.4)	34.3 _(4.6)	33.5 _(4.4)
Generic	34.2 _(5.5)	33.8 _(4.9)	34.0 _(4.9)	33.8 _(4.8)	33.8 _(5.0)	33.6 _(4.9)

(v) Average values and dispersion of Shift, $\langle \theta_4 \rangle \pm (\sigma_{\theta_4})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	0.00 _(0.87)	0.00 _(0.79)	0.00 _(0.79)	0.00 _(0.75)	0.00 _(0.79)	0.00 _(0.75)
CA	0.09 _(0.55)	-0.05 _(0.67)	-0.08 _(0.67)	-0.15 _(0.67)	-0.08 _(0.67)	-0.15 _(0.67)
TA	0.00 _(0.52)	0.00 _(0.58)	0.00 _(0.58)	0.00 _(0.55)	0.00 _(0.58)	0.00 _(0.54)
AG	0.09 _(0.69)	0.19 _(0.64)	0.20 _(0.65)	0.20 _(0.65)	0.20 _(0.65)	0.20 _(0.65)
GG	0.05 _(0.76)	0.02 _(0.67)	0.01 _(0.68)	0.02 _(0.73)	0.01 _(0.68)	0.02 _(0.73)
AA	-0.03 _(0.57)	0.08 _(0.35)	0.08 _(0.35)	0.08 _(0.35)	0.10 _(0.35)	0.10 _(0.35)
GA	-0.28 _(0.46)	-0.21 _(0.49)	-0.22 _(0.49)	-0.21 _(0.50)	-0.22 _(0.49)	-0.21 _(0.50)
AT	0.00 _(0.57)	0.00 _(0.54)	0.00 _(0.54)	0.00 _(0.55)	0.00 _(0.55)	0.00 _(0.55)
AC	0.13 _(0.59)	0.22 _(0.55)	0.22 _(0.55)	0.21 _(0.57)	0.22 _(0.55)	0.21 _(0.57)
GC	0.00 _(0.61)	0.00 _(0.70)	0.00 _(0.72)	0.00 _(0.74)	0.00 _(0.72)	0.00 _(0.74)
Generic	0.00 _(0.64)	0.00 _(0.61)	0.00 _(0.61)	0.00 _(0.62)	0.00 _(0.61)	0.00 _(0.62)

(vi) Average values and dispersion of Slide, $\langle \theta_5 \rangle \pm (\sigma_{\theta_5})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	0.41 _(0.56)	0.36 _(0.57)	0.36 _(0.57)	0.55 _(0.40)	0.36 _(0.57)	0.55 _(0.40)
CA	0.53 _(0.89)	0.22 _(0.91)	0.33 _(0.81)	0.50 _(0.75)	0.33 _(0.81)	0.50 _(0.75)
TA	0.05 _(0.71)	0.37 _(0.93)	0.39 _(0.92)	0.46 _(0.93)	0.43 _(0.95)	0.50 _(0.95)
AG	-0.25 _(0.41)	-0.32 _(0.45)	-0.27 _(0.41)	-0.22 _(0.40)	-0.27 _(0.41)	-0.22 _(0.40)
GG	-0.22 _(0.64)	-0.52 _(0.62)	-0.45 _(0.54)	-0.27 _(0.50)	-0.45 _(0.54)	-0.27 _(0.50)
AA	-0.08 _(0.45)	-0.16 _(0.33)	-0.16 _(0.33)	-0.15 _(0.33)	-0.11 _(0.38)	-0.11 _(0.38)
GA	0.09 _(0.70)	-0.12 _(0.53)	-0.10 _(0.51)	-0.02 _(0.48)	-0.10 _(0.51)	-0.02 _(0.48)
AT	-0.59 _(0.31)	-0.66 _(0.33)	-0.64 _(0.32)	-0.58 _(0.28)	-0.64 _(0.32)	-0.58 _(0.28)
AC	-0.58 _(0.41)	-0.63 _(0.32)	-0.62 _(0.32)	-0.58 _(0.37)	-0.62 _(0.32)	-0.58 _(0.37)
GC	-0.38 _(0.56)	-0.33 _(0.55)	-0.21 _(0.41)	-0.15 _(0.39)	-0.21 _(0.41)	-0.15 _(0.39)
Generic	-0.09 _(0.69)	-0.21 _(0.67)	-0.17 _(0.64)	-0.07 _(0.63)	-0.16 _(0.65)	-0.07 _(0.64)

(vii) Average values and dispersion of Rise, $\langle \theta_6 \rangle \pm (\sigma_{\theta_6})$, in protein-bound DNA dimers

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	3.39 _(0.27)	3.41 _(0.22)	3.41 _(0.22)	3.38 _(0.23)	3.41 _(0.22)	3.38 _(0.23)
CA	3.33 _(0.26)	3.38 _(0.26)	3.37 _(0.23)	3.35 _(0.22)	3.37 _(0.23)	3.35 _(0.22)
TA	3.42 _(0.24)	3.32 _(0.21)	3.33 _(0.21)	3.30 _(0.19)	3.33 _(0.21)	3.31 _(0.19)
AG	3.34 _(0.23)	3.35 _(0.23)	3.35 _(0.23)	3.35 _(0.23)	3.35 _(0.23)	3.35 _(0.23)
GG	3.42 _(0.24)	3.45 _(0.24)	3.45 _(0.24)	3.38 _(0.20)	3.45 _(0.24)	3.38 _(0.20)
AA	3.27 _(0.22)	3.25 _(0.17)	3.25 _(0.17)	3.25 _(0.16)	3.26 _(0.17)	3.25 _(0.16)
GA	3.37 _(0.26)	3.32 _(0.20)	3.32 _(0.21)	3.27 _(0.17)	3.32 _(0.21)	3.27 _(0.17)
AT	3.31 _(0.21)	3.24 _(0.17)	3.24 _(0.17)	3.24 _(0.17)	3.24 _(0.16)	3.24 _(0.17)
AC	3.36 _(0.23)	3.27 _(0.20)	3.27 _(0.20)	3.27 _(0.21)	3.27 _(0.20)	3.27 _(0.21)
GC	3.40 _(0.24)	3.36 _(0.24)	3.33 _(0.24)	3.29 _(0.22)	3.33 _(0.24)	3.29 _(0.22)
Generic	3.36 _(0.25)	3.34 _(0.23)	3.33 _(0.22)	3.31 _(0.21)	3.33 _(0.22)	3.31 _(0.21)

(viii) Relative base-pair flexibility V of protein-bound DNA dimers[†]

Step	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
CG	12.1	4.9	4.9	2.3	4.9	2.3
CA	9.8	7.1	5.1	4.7	5.1	4.7
TA	6.3	7.6	7.5	6.1	8.4	6.8
AG	2.1	1.6	1.3	1.1	1.3	1.1
GG	6.1	4.1	3.6	2.6	3.6	2.6
AA	2.9	0.6	0.6	0.6	0.8	0.8
GA	4.5	1.9	1.8	1.0	1.8	1.0
AT	1.6	0.9	0.8	0.7	0.9	0.7
AC	2.3	0.9	0.9	1.1	0.9	1.1
GC	4.0	3.5	2.2	1.8	2.2	1.8
Generic	9.2	5.2	4.8	4.2	5.1	4.5

[†]Deformability values, in units of ($\text{\AA}^{-\circ}$)³ given by the products of the eigenvalues of the covariance matrix of the averages and products of the six step parameters.

*W.K. Olson, A.A. Gorin, X.-J. Lu, L.M. Hock & V.B. Zhurkin (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci., USA* **95**, 11163-11168.

Table S4. Comparative ‘cost’ of nucleosomal deformation of individual base-pair steps with knowledge-based functions based on different subsets of observed protein-bound DNA conformations[†]

(i) Pyrimidine-purine (YR) base-pair steps

CG

Score	Olson <i>et al.</i> [*]	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	7.4	4.5	4.5	6.4	4.5	6.4
E_0	0.8	0.4	0.4	0.7	0.4	0.7
E^\dagger	24.8	13.0	13.0	18.8	13.0	18.8
SH_0	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4
SH^\dagger	-1.7	1.5	1.5	1.5	1.5	1.5

CA

Score	Olson <i>et al.</i> [*]	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	8.6	4.0	4.4	4.4	4.4	4.4
E_0	1.3	0.4	0.4	0.6	0.4	0.6
E^\dagger	30.6	13.8	14.9	14.6	14.9	14.6
SH_0	-2.4	0.1	-1.0	-1.0	-1.0	-1.0
SH^\dagger	-2.0	-2.0	-2.0	-1.5	-2.0	-1.5

Pyrimidine-purine (YR) base-pair steps (continued)

TG

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	8.2	3.8	4.2	4.2	4.2	4.2
E_0	1.0	0.4	0.5	0.7	0.5	0.7
E^\dagger	29.1	13.1	14.5	14.3	14.5	14.3
SH_0	0.1	-2.4	-2.4	-1.0	-2.4	-1.0
SH^\dagger	-1.7	-1.7	-1.7	-1.9	-1.7	-1.9

TA

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	8.7	4.3	4.3	4.7	4.6	4.8
E_0	1.0	0.6	0.7	0.8	0.7	0.8
E^\dagger	26.6	14.8	14.8	19.1	16.9	19.4
SH_0	-2.4	0.1	0.1	0.1	0.1	0.1
SH^\dagger	-1.9	-1.5	-1.5	-1.5	-1.5	-1.5

(ii) Purine-purine and pyrimidine-pyrimidine (RR, YY) base-pair steps*AG*

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	18.1	9.4	10.7	11.1	10.7	11.1
E_0	1.0	0.8	0.8	0.9	0.8	0.9
E^\dagger	67.1	33.4	41.8	44.0	41.9	44.0
SH_0	1.0	-0.1	-0.1	-0.1	-0.1	-0.1
SH^\dagger	1.5	1.5	1.5	1.5	1.5	1.5

CT

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	19.4	10.2	11.5	12.0	11.5	12.0
E_0	0.6	0.2	0.2	0.2	0.2	0.2
E^\dagger	59.4	31.4	39.0	40.7	39.0	40.7
SH_0	0.1	0.1	0.1	0.1	0.1	0.1
SH^\dagger	1.5	1.5	1.5	1.5	1.5	1.5

Purine-purine and pyrimidine-pyrimidine (RR, YY) base-pair steps (continued)

GG

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	10.3	5.4	5.9	5.7	5.9	5.7
E_0	0.9	0.5	0.4	0.3	0.4	0.3
E^\dagger	44.4	17.9	18.5	18.9	18.4	18.9
SH_0	0.1	0.1	0.1	0.1	0.1	0.1
SH^\dagger	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0

CC

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	11.4	6.0	6.5	6.6	6.5	6.6
E_0	0.7	0.5	0.5	0.3	0.5	0.3
E^\dagger	39.0	18.2	18.9	17.8	18.9	17.8
SH_0	0.1	0.1	0.1	0.1	0.1	0.1
SH^\dagger	-2.0	-2.0	-2.0	-1.9	-2.0	-1.9

Purine-purine and pyrimidine-pyrimidine (RR, YY) base-pair steps (continued)

GA

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	12.3	7.5	7.5	8.7	7.5	8.7
E_0	1.0	0.7	0.7	0.9	0.7	0.9
E^\dagger	47.3	24.6	24.4	31.7	24.4	31.7
SH_0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
SH^\dagger	-1.7	-1.5	-1.5	-1.5	-1.5	-1.5

TC

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	13.5	8.0	8.1	9.4	8.1	9.4
E_0	2.0	1.1	1.2	1.3	1.2	1.3
E^\dagger	45.7	25.7	25.8	30.4	25.8	30.4
SH_0	-0.9	0.1	0.1	-0.9	0.1	-0.9
SH^\dagger	2.5	-1.6	-1.6	-1.6	-1.6	-1.6

Purine-pyrimidine (RY) base-pair steps (continued)

GT

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	20.3	15.6	15.6	12.5	15.6	12.5
E_0	1.1	0.9	0.9	0.7	0.9	0.7
E^\dagger	62.4	44.1	44.1	38.4	44.1	38.4
SH ₀	1.2	1.2	1.2	1.2	1.2	1.2
SH [†]	-2.0	1.5	1.5	-2.0	1.5	-2.0

GC

Score	Olson <i>et al.</i> *	A+B+AB	B+AB	B	B+AB+TA	B+TA
$\langle E \rangle$	15.7	7.1	8.8	7.1	8.8	7.1
E_0	1.9	1.0	1.1	-1.9	1.1	-1.9
E^\dagger	89.8	39.8	47.1	32.0	47.1	32.0
SH ₀	1.2	1.2	1.2	-2.3	1.2	-2.3
SH [†]	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0

† Knowledge-based scores — average cost $\langle E \rangle$ over the central 60 base-pair steps of the best-resolved nucleosome core particle structure (2) and the least and most costly values, E_0 and E^\dagger — derived from the base-pair step parameters of DNA dimers of different conformational types from high-resolution protein-DNA structures. Locations SH₀ and SH[†] of the least and most costly steps are expressed in terms of superhelical position, *i.e.*, number of helical turns with respect to the structural dyad. See text and legend to Table 1

*W.K. Olson, A.A. Gorin, X.-J. Lu, L.M. Hock & V.B. Zhurkin (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, Proc. Natl. Acad. Sci., USA 95, 11163-11168.

Table S5. Threading scores of known nucleosome binding and antiselection sequences on the central 60 base-pair steps of the 147-bp nucleosome core-particle structure (NDB_ID pd0287) (2) and an ideal 61-bp superhelical template with the same average global structure.[§]

Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
Nucleosome-binding sequences from the mouse genome								
<i>A-tracts 1</i>								
phcn4	501	43	415	633	146	3	140	152
phcn8	511	42	431	618	147	3	142	153
phcn12	510	42	408	632	146	4	139	158
phcn14	507	41	421	649	146	3	141	154
phcn17	511	42	430	594	145	3	139	152
phcn18	500	51	396	629	148	3	140	154
phcn21	499	45	393	638	145	3	138	152
phcn22	495	42	412	615	143	3	137	148
phcn23	498	45	391	635	144	4	139	155
phcn24	493	45	382	620	141	2	137	146
phcn26	499	45	387	637	144	2	140	149
phcn29	501	47	378	640	144	3	138	150
phcn39	494	44	400	640	144	3	137	150
phcn43	490	37	423	625	141	3	135	149
phcn46	502	41	414	618	146	4	136	154
phcn47	506	41	409	619	145	3	137	152
phcn50	494	40	406	607	142	2	138	147
phwn12	501	41	408	592	146	2	143	152
<i>A-tracts 2</i>								
phcn2	503	44	408	598	144	2	137	148
phcn10	499	41	401	604	143	3	136	148
phcn13	491	45	382	635	145	3	139	151
phcn20	506	43	410	593	146	3	139	151
phcn30	494	40	397	597	144	5	133	153
phcn42	509	43	416	601	146	3	141	153
phwn8	517	45	418	606	145	3	140	150
phwn26	500	45	410	607	145	3	140	152

Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
phwn27	511	44	410	595	145	3	138	152
<i>A-tracts 3</i>								
phcn3	501	39	393	607	144	3	136	150
phcn7	493	47	381	614	143	2	139	148
phcn9	510	43	403	597	146	5	135	154
phcn11	510	49	414	616	144	4	137	151
phcn19	494	43	399	587	143	3	136	149
phcn28	500	39	393	600	144	2	141	149
phcn37	497	46	389	638	144	2	139	150
phcn44	500	46	387	636	143	2	139	149
phwn11	496	42	386	615	143	3	137	148
phwn29	509	44	401	597	146	5	135	153
phwn36	513	41	419	596	146	4	137	153
<i>TG/CTG runs</i>								
phcn31	531	40	446	619	161	5	150	170
phcn32	535	58	417	623	163	4	151	171
phcn33	550	47	452	621	168	4	161	176
phcn40	527	32	458	642	155	6	144	168
phwn2	496	42	416	589	148	4	140	158
phwn3	518	36	449	622	151	5	140	162
phwn4	498	42	421	590	149	3	141	157
phwn7	461	33	380	555	137	5	128	148
phwn10	525	50	427	609	154	3	147	162
phwn13	536	58	414	644	161	3	156	167
phwn14	527	51	381	611	156	7	136	166
phwn15	560	50	457	656	169	2	163	176
phwn16	499	43	422	614	145	5	134	155
phwn17	502	39	396	584	150	6	136	163
phwn18	495	50	401	602	144	4	134	152
phwn20	495	41	409	589	143	3	135	151
phwn22	508	33	430	595	151	5	141	162
phwn23	528	43	427	650	159	4	152	169
phwn24	507	37	432	615	147	6	136	162

Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
phwn25	492	44	385	581	145	6	132	157
phwn28	506	36	433	621	146	6	136	158
phwn30	495	36	409	597	147	4	138	157
phwn32	479	38	407	561	146	7	132	156
<i>Phased TATA</i>								
phcn5	502	38	436	589	146	13	127	173
phcn25	513	36	431	582	147	11	131	170
phcn41	509	38	415	583	149	12	131	171
phcn49	520	41	430	610	152	12	133	174
<i>Phased TG/CA</i>								
phcn6	528	43	434	636	156	3	147	163
phcn16	523	47	405	622	155	3	147	159
phcn27	530	46	417	628	156	2	150	160
phcn48	529	47	447	642	154	2	148	158
<i>No sequence</i>								
phcn1	483	33	407	567	141	4	132	149
phcn15	492	39	420	574	142	5	129	152
phcn34	511	39	436	598	147	5	137	157
phcn35	467	35	371	581	133	5	123	146
phcn36	493	41	385	597	144	7	123	155
phcn38	467	35	371	581	133	5	123	146
phcn45	470	38	382	593	134	7	121	148
phwn1	467	32	383	535	136	6	123	149
phwn5	491	43	410	582	141	5	132	151
phwn6	465	34	387	540	135	7	119	155
phwn9	473	40	397	566	136	4	127	146
phwn19	498	38	420	603	144	4	137	152
phwn21	477	39	379	562	134	5	120	145
phwn31	494	46	386	597	142	5	132	152
phwn33	471	38	394	584	141	8	126	158
phwn34	503	39	385	595	143	5	134	153
phwn35	467	36	386	560	131	8	116	147
phwn37	450	27	393	528	133	5	121	143

Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
Anti-selection sequences from the mouse genome								
<i>TGGA fragments</i>								
35	500	25	442	541	154	3	148	158
19	492	35	431	576	153	2	148	156
29	480	39	397	566	149	3	143	156
75	495	50	410	600	152	2	149	158
44	486	34	413	567	146	3	140	154
81	482	35	416	569	149	3	143	154
49	474	31	426	528	147	3	143	152
62	474	44	374	566	148	2	144	152
57	472	32	408	548	147	2	143	152
47	480	41	369	571	146	2	140	150
77	481	40	404	566	149	3	142	155
23	478	42	397	572	146	3	140	152
80	482	31	433	557	149	2	146	153
86	463	43	371	566	144	2	140	151
<i>Badsecs</i>								
24	490	31	424	562	144	5	133	154
26	498	33	446	569	142	1	140	145
28	511	51	418	653	148	3	143	152
48	502	36	420	582	145	6	134	160
51	500	33	418	564	140	4	133	148
78	503	34	397	556	146	6	134	157
31	508	41	432	571	147	3	140	152
32	516	49	427	627	149	3	140	154
33	518	38	424	597	149	8	129	163
52	494	30	428	551	143	6	133	153
63	526	36	433	595	147	7	138	161
69	508	37	454	615	146	4	131	153
73	506	35	428	589	146	4	139	156
76	493	34	430	593	144	4	135	152
84	486	27	433	544	144	6	134	157
85	503	34	418	594	149	4	141	157

Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
27	452	38	386	532	128	4	119	138
30	490	38	395	559	142	4	134	148
43	484	35	389	579	136	4	130	142
53	488	30	437	549	140	4	132	147
55	489	36	422	556	142	3	135	146
56	493	28	450	538	142	6	133	149
59	485	49	397	593	142	4	134	151
66	515	42	397	616	147	5	138	156
67	521	48	444	602	152	4	141	158
74	512	40	432	602	143	4	137	151
Synthetic SELEX sequences								
10nvp	477	32	407	561	136	5	121	150
11nvp	483	41	390	579	137	5	126	149
13nvp	467	36	368	615	132	7	114	146
16nvp	488	36	371	595	139	6	124	151
18nvp	479	40	362	611	138	5	127	149
22nvp	480	38	376	577	138	6	124	153
24nvp	481	34	364	575	139	5	122	148
25nvp	479	41	368	629	139	5	128	152
26nvp	493	38	397	589	141	7	124	155
28nvp	485	40	368	597	140	7	124	155
29nvp	484	42	390	609	140	6	124	150
2nvp	490	38	400	630	140	4	130	147
30nvp	481	31	416	572	138	5	126	154
34nvp	483	41	345	584	138	4	128	148
37nvp	471	38	359	586	138	5	126	149
38nvp	471	42	382	605	136	5	122	148
40nvp	469	41	345	601	134	6	122	152
41nvp	477	37	399	562	136	4	125	150
52nvp	481	38	378	589	137	5	123	146
54nvp	484	41	393	643	138	7	123	155
5nvp	483	38	394	599	140	4	128	151
602nvp_rev_rev	477	37	382	573	136	7	124	154

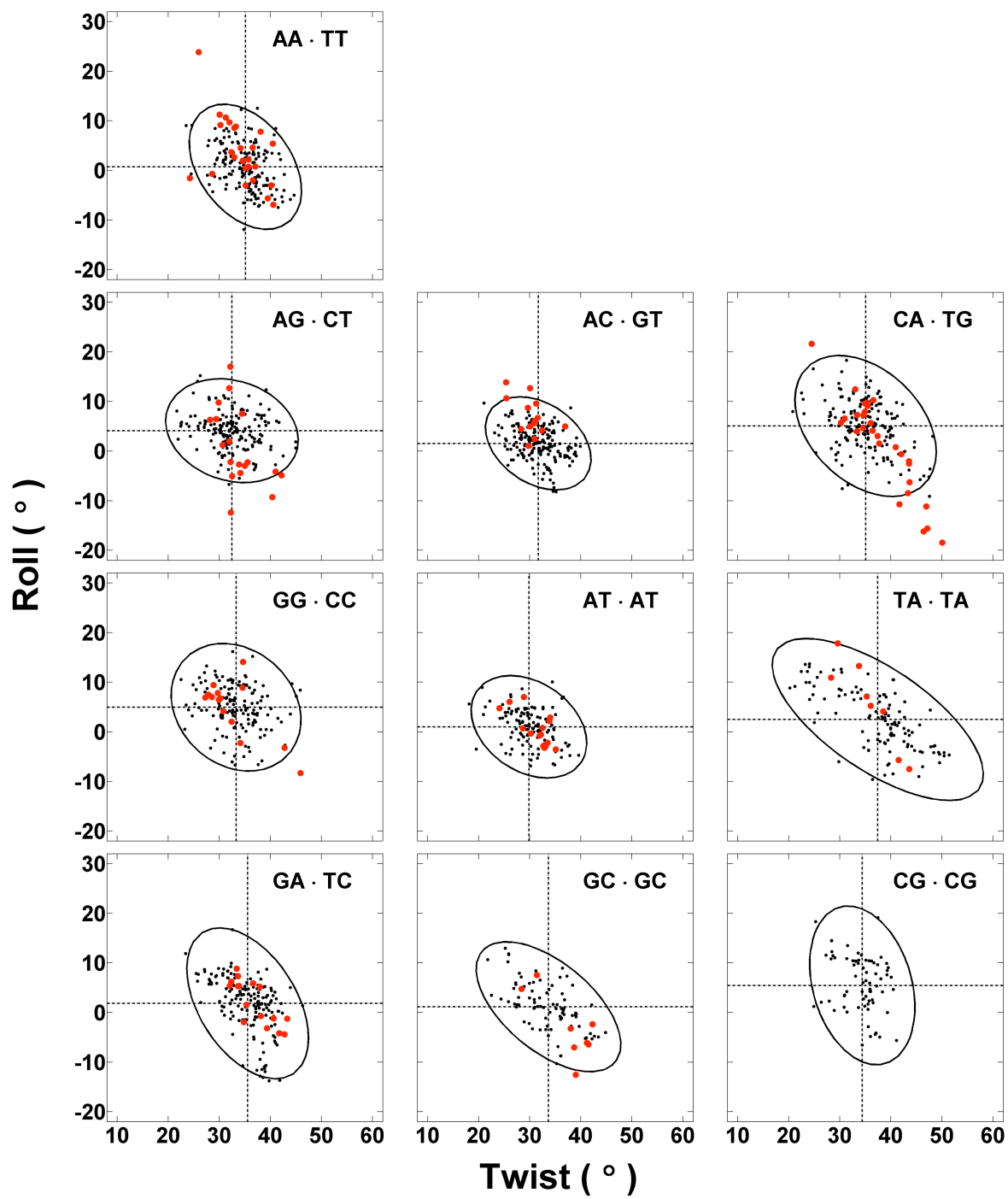
Sequence_ID	<i>Crystallographic template</i>				<i>Smooth superhelical template</i>			
	$\langle U \rangle$	σ_U	U_0	U^\dagger	$\langle U \rangle$	σ_U	U_0	U^\dagger
603nvp	487	44	358	618	138	5	124	150
604rnvp	496	37	415	594	141	5	129	154
605nvp	489	39	371	575	139	5	126	153
607nvp_rev_rev	476	39	347	571	135	4	123	146
609nvp	475	41	381	587	136	7	118	152
611nvp	466	37	361	585	132	5	121	143
613nvp	476	41	372	583	138	5	126	149
615fnvp	475	39	342	590	134	6	123	152
618nvp_rev_rev	479	37	390	587	137	4	128	146
620nvp_rev_rev	470	40	366	577	133	5	121	146
621rnvp	480	36	390	565	137	5	125	152
624bnvp	481	36	389	570	137	4	129	145
626nvp	488	34	399	579	139	5	126	150
628nvp_rev_rev	477	43	374	586	136	8	117	157
69nvp	475	44	373	610	135	7	119	149
6nvp	474	39	386	576	134	5	122	149
73nvp	485	44	358	628	139	5	123	148
7nvp	485	35	408	593	137	5	128	149
83nvp	472	42	340	630	137	6	123	152

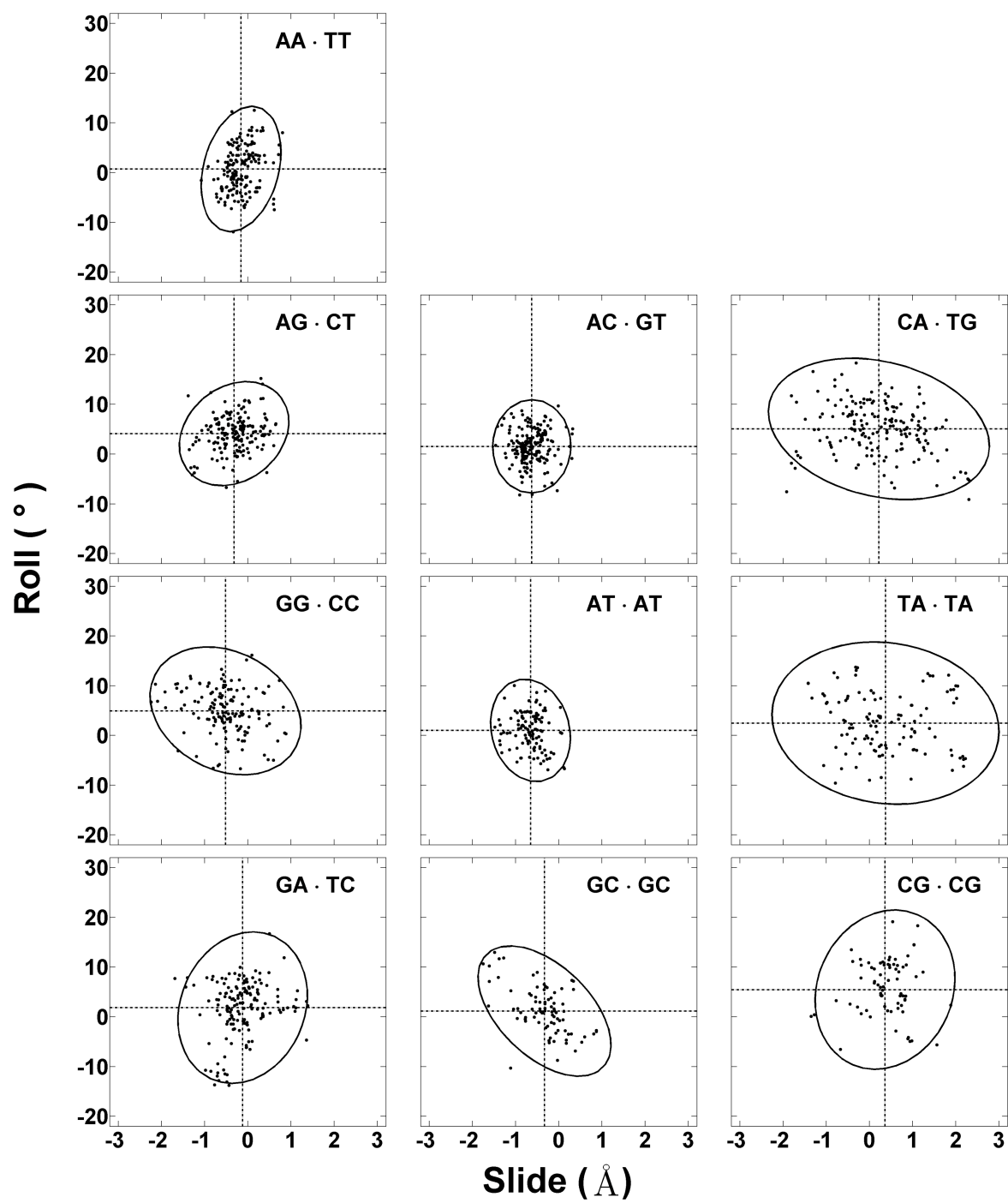
[§] Threading scores reported for individual sequences are the total dimeric scores computed with Eqn. (1) using all six step parameters and elastic constants and rest states based on the dimer steps from all conformational categories, *i.e.*, the A+B+AB dataset in Table 1. Values include the mean score $\langle U \rangle$ and the standard deviation σ_U for all possible settings of each sequence on the designated templates as well as the lowest and highest scores U_0 and U^\dagger associated with the best and worst settings of each sequence on the template.

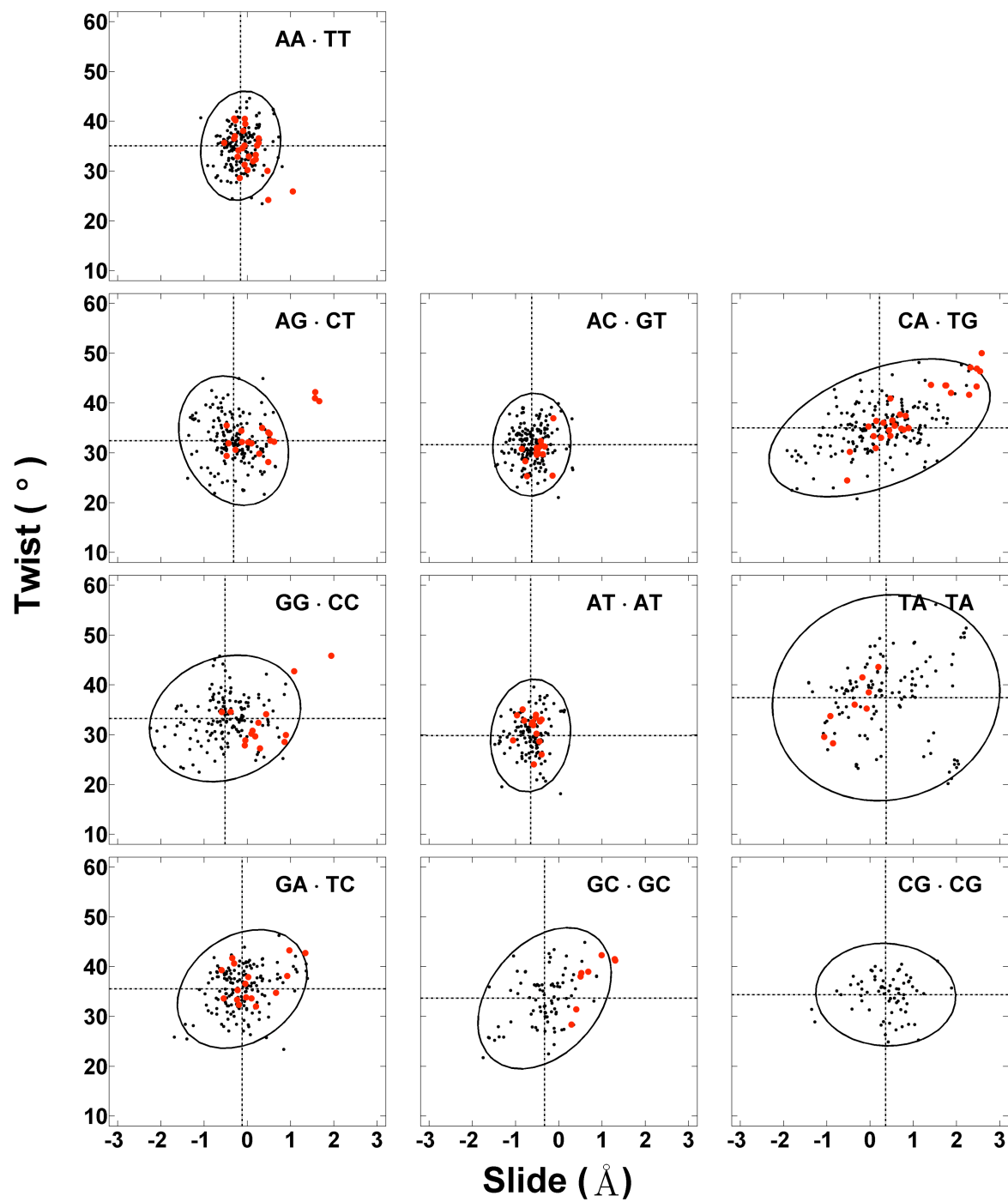
Legends to Supplementary Figures

Figure S1. Collective scatter plots of the rigid-body (step) parameters found for the ten unique DNA base-pair steps in the high-resolution protein-DNA crystal complexes that make up the reference database, the sequence-dependent potentials derived from these data, and the step-parameter values found in nucleosomal DNA. Small black dots correspond to the points used to derive the potentials (solid contours) and large red dots to the states that the given dimers adopt in the currently best-resolved nucleosome core particle structure (NDB_ID pd0287, Davey *et al.* 2002. *J. Mol. Biol.* **319**, 1097-1113). Ellipses are projections of the multi-dimensional potentials on the specified conformational planes obtained from the 2×2 covariance matrices of observed values: (a) Roll-Twist; (b) Roll-Slide; (c) Twist-Slide. Contours correspond to deviations of parameters equal to two times the combined root-mean-square deviations from the intrinsic (mean) values. Average values of step parameters are highlighted by thin (dashed) lines. The three columns show the respective deformational patterns of individual purine-purine (RR), purine-pyrimidine (RY), and pyrimidine-purine (YR) steps. Note the absence of CG steps in the nucleosomal DNA. Images kindly provided by Guohui Zheng.

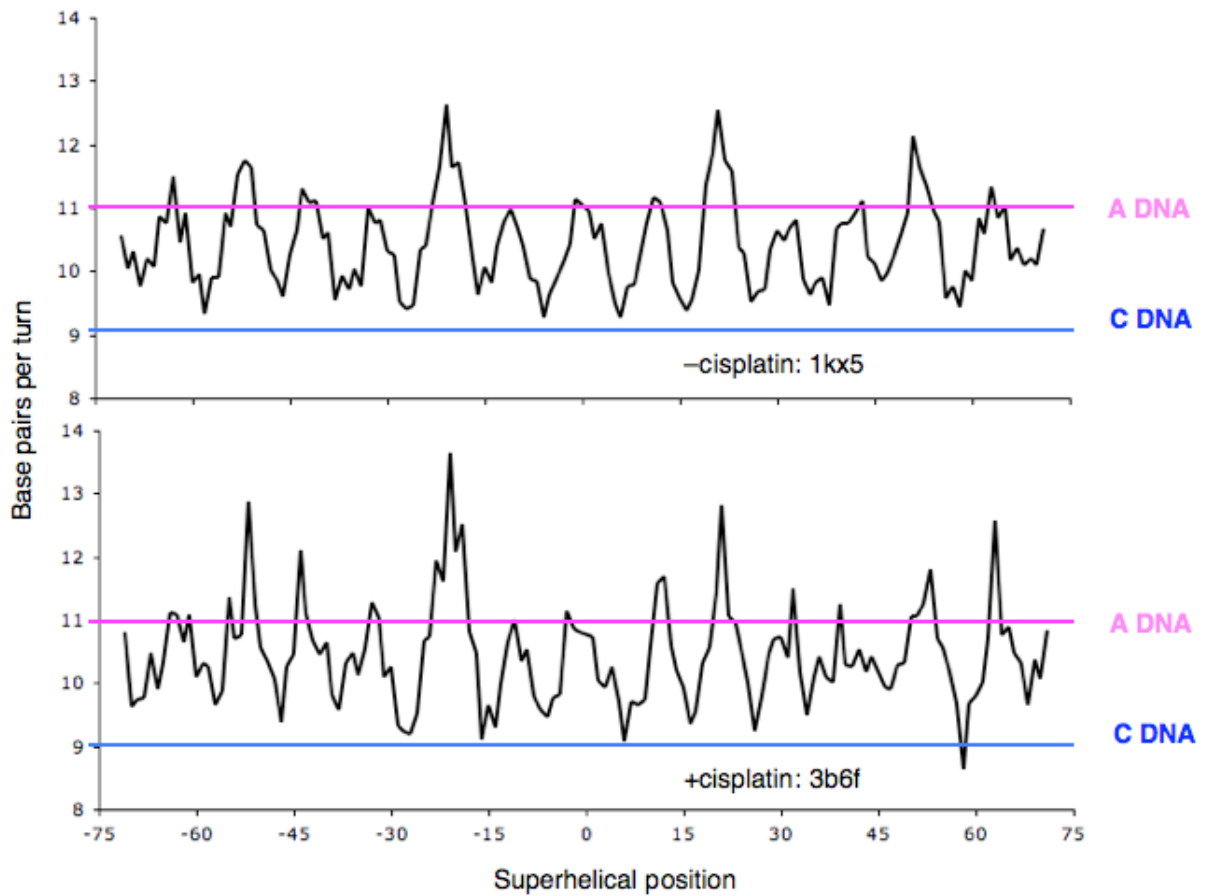
Figure S2. Comparative variation of Twist, expressed the number of base pairs that would form a complete turn of an ideal, naturally straight helix, as a function of DNA superhelical position in two 147-bp nucleosome core-particle structures: (top) the DNA pathway in the currently best-resolved nucleosome core particle structure (NDB_ID pd0287/PDB_ID 1kx5, Davey *et al.* 2002. *J. Mol. Biol.* **319**, 1097-1113); (bottom) the pathway adopted by the same sequence in the presence of cisplatin (NDB_ID pd1046/PDB_ID 3b6f, Wu & Davey. 2008. *Nature Chem. Biol.* **4**, 110-112). Superhelical positions correspond to the number of double-helical turns a dimeric step is displaced from the structural dyad on the central base pair (here denoted by 0). Color-coded lines denote the helical repeat in the canonical A-DNA (Franklin & Gosling. 1953. *Nature* **171**, 740-741) and C-DNA (Marvin *et al.* 1958. *Nature* **182**, 387-388) double helices.

Balasubramanian *et al.*, Figure S1a

Balasubramanian *et al.*, Figure S1b

Balasubramanian *et al.*, Figure S1c

The presence of cisplatin appears to unwind nucleosomal DNA to more A-like states.



Balasubramanian *et al.*, Figure S2